YOU CAN HEAR THE SHAPE OF A ROOM: ACOUSTIC EXPLORATION AND RECONSTRUCTION OF CONVEX POLYHEDRA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mingbo Zhao

January 2012

YOU CAN HEAR THE SHAPE OF A ROOM: ACOUSTIC EXPLORATION AND RECONSTRUCTION OF CONVEX POLYHEDRA

Mingbo Zhao, Ph.D.

Cornell University 2012

The Thesis: Radar/sonar/lidar are very well-studied means for the detection of objects that are distant from the observer and for the estimation of such physical properties of the object as its distance from the observer, surface reflectivity, rotation, and velocity. This dissertation focuses instead on the novel issue of estimating the surface of a three-dimensional convex room that is empty except for the acoustic-based measurement system introduced by the observer. While acoustic sounding is the essence of sonar, our objective of characterizing a convex polyhedral room is unlike any exploration undertaken hitherto.

Approach and Results: We first deploy a single omnidirectional (its gain pattern need not be directionally uniform but is always positive) sound source at a location of our choosing that will be taken to be the origin of our coordinate system. This source will be controlled to emit a short pulse of duration T^* and known signal shape $s(t)$. Subsequent pulses may be generated provided the interpulse interval is sufficiently long.

We then deploy an array of omnidirectional (again, they can have direction-dependent gains that are always positive) microphones at known locations. We assume that the walls of the room are not only planar but also have surfaces that yield specular acoustic reflection that is akin to the reflection of light from a mirror. Each microphone is monitored to record both the direct line-of-sight pulse from the source and a first received echo. The records of the originating pulse

and the two pulses recorded at each microphone are then processed centrally to infer the placement of the walls of the room.

As we do not know the true number W of walls, we cannot guarantee that we will detect each of them. As we shall see, it takes four microphones receiving a first echo from the same wall to enable us to locate that wall. Hence, we would need a minimum of $4W$ microphones to ensure detection of all walls. As W is unknown, the detection of all walls cannot be guaranteed. Hence, our approach is only known to generate a convex polyhedral upper bound to the true convex polyhedral room.

We will treat both the case of noiseless reception at the microphones and the more complex but more realistic case of noisy reception. The theoretical results achieved are sometimes supplemented by simulated examples and/or by a laboratory experiment.

BIOGRAPHICAL SKETCH

My name is Mingbo Zhao, and I am currently working in Lutron Electronics Inc. as a Design and Development Engineer.

I was born in Tianjin, China on Feb 12, 1981, to Jun Zhao and Yanfen Zhou. In 1999, I moved to Philadelphia, PA, United States, and attended Drexel University after my graduation from Tianjin Nankai High School. I studied Computer Engineering in Drexel University for 2 years and then transferred to Cornell University, Ithaca, NY. I obtained my Bachelor of Science degree in Electrical and Computer Engineering in 2003 and my Master of Engineering degree in 2004. In 2004, I joined the MS/PhD program in Cornell Electrical and Computer Engineering department, where I conducted research in the Communications Network Group supervised by Prof. Sergio Servetto. My MS/PhD research work focused on developing algorithms and experiments to acoustically sense the shape of an unknown room. My Special Committee was composed of Assistant Professor Sergio Servetto as the Chair, Prof. Michael J. Todd (Cornell ORIE department), and Prof. Alexander Vladirmirsky (Cornell Mathematics department). On July 24, 2007, Prof. Sergio Servetto died in an airplane crash. After this tragic accident, I changed my advisor and Special Committee chair to Prof. Terrence Fine and maintained the same research focus. In January 2009 I took a leave of absence and joined Lutron Electronics Inc., where I was tasked with designing large-scale commercial and residential lighting control systems.

This study is dedicated to my mother, Yanfen Zhou and my father, Jun Zhao whose love, encouragement, support and inspiration have made all of this possible.

ACKNOWLEDGEMENTS

I would like to thank first my advisor, Prof. Terrence Fine. His constant support and encouragement not only made this dissertation possible but also influenced my whole life. He has instilled in me, by example, a strong sense of discipline and integrity for which I am very much grateful. I am truly blessed in having Prof. Terrence Fine as my advisor. Additionally, I would like to thank my PhD committee members, Prof. Michael J. Todd and Prof. Alexander Vladimirsky. Their support and insights made this dissertation much stronger.

I would also like to thank my parents for their unconditional support. I am grateful to my wife, Qiuyue Yang, for her love and sacrifice. Without her, this dissertation would not be possible. Baokang and Baobao, my kids, you two are my constant inspirations to strive for more. I always think of you. Additionally, I would like to thank Prof. Jiuming Li of Nankai University, Tianjin, China, for making my transition to the United States possible. Finally, I would like to thank Prof. Sergio Servetto for initiating my PhD research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

MOTIVATION, MODELS, AND PROBLEM FORMULATION

1.1 Motivation and Approach

1.1.1 Motivation

In August 2007, a series of mine crashes in Huntington, Utah claimed nine miners' lives. Three of those nine were the rescuers who were searching for the six buried during the first crash. Prior to the insertion of the three rescuers, some "smart" equipment, such as video cameras, were sent underground to locate those trapped. However, this equipment could not be configured to function well in that disastrous environment. The smart tools were unable to navigate in the coal mine miles below the surface. As a result, human beings were sent to physically locate and rescue the miners, and this led to the loss of additional lives. A similar need for mapping an interior space can arise in a military setting when there may be enemy combatants or booby traps hidden in a structure of unknown configuration.

We propose to map the interior of a "room" by using the specular reflection of acoustic waves that are generated by a single sound source and recorded by an array of microphones. Alternative physical modalities include illuminating the interior with beams of optical or Terahertz wavelengths. Using either of these modalities we can hope to recover the shape of the interior. For definiteness and the ease of implementing laboratory experiments, this dissertation focuses on using acoustic waves.

1.1.2 Approach

- We assume that our "room" is restricted to be a convex polyhedron having an unknown number of faces and unknown geometry. There are no preferred directions.
- An omnidirectional (its positive gain can vary with direction) sound source establishes a sound field in the interior of the region being examined.
- A seemingly similar question (e.g., see Mark Kac [11]) about hearing the shape of a drum, expressed in terms of an eigenfunction spectrum, was answered negatively (see Gordon, Webb, Wolpert [7])).
- We avoid this inability to determine shape by using the transient behavior of our acoustic field produced by short pulses from a single sound source, and show that one can learn the shape of the room containing the field, at least under certain assumptions.
- An array of omnidirectional (again, their gain can vary with direction) microphones is deployed to make local measurements of the transient spatial acoustic wave contained in the convex polyhedron under consideration.
- Our goal is to recover a description of the convex polyhedron itself from these local wave measurements.
- The collection of microphone measurements is subject to corruption by ambient noise and noise internal to the microphone. Ambient noise might be produced by other wave sources (e.g., wind).
- When considering a sensor network another fundamental source of noise in the measurements collected by this device is random phase variations

across nodes, derived from inherently imperfect time synchronization methods.

- We will treat both the noiseless and noisy cases.

1.2 Definitions and System Model

1.2.1 Definitions

The following definitions and terminology will be used throughout this dissertation.

- All of our considerations are focussed on the three-dimensional case of $\mathbb{R}^3$, although reference is sometimes made to the two-dimensional case for clarity in an example.
- A *half-space* H is determined by an oriented plane with normal vector $\mathbf{I}$ and point on the plane $\mathbf{b}$ as follows

  $$x, \mathbf{I}, \mathbf{b} \in \mathbb{R}^3 \text{ and } H = \{x : \mathbf{I} \cdot (x - \mathbf{b}) \leq 0\}.$$

  Equivalently, for a scalar constant c,

  $$H = \{x : \mathbf{I} \cdot x \leq c\}.$$

- A *convex* polyhedron $\mathcal{P}$ with n faces is a polyhedron defined as the intersection of n half-spaces,

  $$\mathcal{P} = \cap_{i=1}^{n} H_i.$$

1.2.2 Acoustic Wave Model

1. We assume that a planar reflecting wall has a surface that enables *specular reflection*. The walls are mirror-like and reflected sound is neither dispersed nor diffused.
2. In this case, the acoustic reflection satisfies the Laws of Reflection: the angle of incidence is equal to the angle of reflection; and the normal to the reflecting plane, the incident path and the reflected path are co-planar.
3. When the Laws of Reflection are satisfied, then the *reflection* of a point sound source S at the origin $\mathbf{0}$ can be considered as if it is emitted by the same S located at $\mathbf{I}$, lying behind the planar reflecting wall $\mathcal{W}$. The wall $\mathcal{W}$ perpendicularly bisects the line segment from the origin to the image point $\mathbf{I}$. Equivalently, $\mathbf{I}$ is a normal to $\mathcal{W}$ and lies at distance $\frac{1}{2}|\mathbf{I}|$ from the origin (see Allen and Berkley [1] for image points).

1.2.3 Physical Model

We assume the following physical model descriptions throughout this dissertation.

1. The environment is a finite-volume subset of $\mathbb{R}^3$. The case of $\mathbb{R}^2$ is less realistic and easier and can be treated similarly.
2. The model for the environment we are trying to identify is a room $\mathcal{R}$ that is a convex polyhedron with an unknown (possibly randomly chosen) number W of specularly reflecting faces or walls and unknown locations of these walls.

3. A *point source* S, is defined in terms of a function of time, $4\pi s(t)$, an angular directional gain function $G(\theta)$, and the speed of sound c in the medium of the interior of the room, and it is such that the signal strength at a location at distance $d > 0$ from the source in direction θ is given by

$$\frac{G(\theta)}{d} s\left(t - \frac{d}{c}\right).$$

For simplicity in what follows we take $G(\theta) = A$, constant gain in all directions. The time shift by d/c is the propagation delay.

4. Again, for simplicity, we assume that the directional gain of microphones is constant in all directions.

5. We assume, without loss of generality, that the point sound source S is located at the origin of our coordinate system and is capable of emitting short duration, high amplitude pulses. The desired pulse duration is determined as part of the setup of the system so that at no microphone is there overlap between the received line-of-sight pulse and the received first reflected pulse (see Section 1.3).

6. The *image point* $\mathbf{I}$ of a point source S *with respect to a face/plane/wall* $\mathcal{W} = \{\mathbf{x} : \mathbf{a_i} \cdot \mathbf{x} = b_i\}$ is the unique point such this plane is the perpendicular bisector of the line segment joining $\mathbf{I}$ and the origin.

7. The image point source S' at location $\mathbf{I}$ has the same physical characteristics, given in Point 3, as the original source S at the origin.

8. The location $\mathbf{I}$ of an image point determines a plane π specified by a normal vector $\mathbf{I}$ and a point $\frac{1}{2}\mathbf{I} \in \pi$ through:

$$\pi = \{\mathbf{x} : \mathbf{I} \cdot \mathbf{x} = \frac{1}{2}||\mathbf{I}||^2\},$$

where $||\mathbf{x}||$ denotes the Euclidean length of $\mathbf{x}$.

9. Given that $\mathbf{I}$ is an acoustic image point, then this plane π must contain a polygonal face of $\mathcal{R}$.

10. Assuming that there are W walls, with W initially unknown to us, we have that these W walls are polygonal subsets of $\{\pi_i : i = 1 : W\}$ determined by $\{\mathbf{I}_i, i = 1 : W\}$, which are initially also unknown to us. Let H_i denote the half-space containing the sound source at $\mathbf{0}$ that is determined by the plane π_i through:

$$H_i = \{\mathbf{x} : \mathbf{I}_i \cdot \mathbf{x} \leq \frac{1}{2}\|\mathbf{I}_i\|^2\}.$$

The intersection $\mathcal{R} = \bigcap_{i=1}^{W} H_i$ determines the desired polyhedron subsets that are the walls or faces of the polyhedral room.

11. The number of walls W is initially unknown, assumed to be at least four, the minimum needed to form a closed room. W also will be modeled in Chapter 3 as a random variable with a Geometric distribution.

12. Our identification of $\mathcal{R}$ will be based upon using echoes to identify the set of image points and then constructing $\mathcal{R}$ from these image points. Should we not locate all of the image points, then the resulting reconstruction $\mathcal{R}_{est}$ will contain $\mathcal{R}$.

13. We assume that we have an array of K *omnidirectional microphones/sensors* located at $M = \{\mathbf{m}_1, \ldots, \mathbf{m}_K\}$ and all lying on the same side of the walls as the sound source at $\mathbf{0}$. We cannot place microphones outside the room we are estimating.

14. The disposition of the microphones in the array may be either chosen as the vertices of some geometric figure or chosen randomly.

15. As we construct M, we do know K. However, for progress in some of our subsequent analysis, we will assume that K is a Poisson random variable (see Subsection 3.3.1).

1.2.4 Experimental Hardware Model Description

A more ambitious plan to verify performance of the algorithms for detecting the shape of rooms was scaled back due to various practicalities such as loss of lab space and budget for equipment. However, the results of some confirming experiments are reported in Chapter 6.

1. We use one set of four speakers facing opposite directions on a horizontal plane to generate pulse sound and an array of microphones to sample the resultant acoustic wave field.
2. We assume that the *speaker set* is an approximate point source that generates omnidirectional acoustic wave that propagates in all directions. It will turn out that our analysis is relatively insensitive to true signal amplitudes.
3. A *microphone* is a device that measures the temporal variations in the intensity of a wave field at a given location, such as samples of the wave field at an approximate point. Each microphone is omnidirectional, capable of receiving signals arriving from any direction. The overall system is a distributed collection of speakers and microphones that are coordinated through a centralized computing unit that delivers the required information. Each speaker has adjustable power and controllable output. Each microphone has its own internal clock.
4. We only use up to the first echo recorded from the microphone samples due to the attenuation of secondary echoes that are transmitted over longer distances than the first echo.
5. We assume that the sound pulse $s(t)$ generated by the sound source has

support $[0, T^*]$, that is, $s(t) = 0$ outside of this support set.

6. We assume that $\|\mathbf{m_i}\|$ is known and the distance between the microphone and the nearest image point, $\|\mathbf{m_i} - \mathbf{I}\|$, is unknown but is large enough so that there is no overlap between the echo and direct waves picked-up at the microphone. This is a critical assumption, as our eventual algorithm depends on the direct wave to estimate the delay to a first echo.

1.2.5 Sources of Noise and Uncertainty

We identified several kinds of noises and uncertainties that limit the accuracy with which we can recover the unknown room. Throughout this dissertation, the noise involved is referred to as $n(t)$, and it includes the following possible components.

1. Ambient noise in the space being surveyed that arises from: sounds of air flow; noise from motors in computer hard drives and fans; characteristics of space being examined such as water flow or animal occupants. The signal measured by microphones under ambient noise is the sum of the desired acoustic wave field and the ambient noise wave field. In the absence of ambient noise, the signals measured by microphones are the desired samples of the wave field.
2. Electro-mechanical noise generated internally by the microphone itself so that its electrical response is not a perfect representation of the incident pressure fluctuations. Typically, such noise is a wide-spectrum Gaussian random process.

3. Limited precision in our knowledge of the acoustic pulse $s(t)$ we generate, whose echoes we will be using to infer the location of the image source point.
4. Phase perturbation or timing errors due to imperfect synchronization: The clock at each microphone differs from the clock at the centralized computing unit; therefore the signals measured by the microphones are subject to random time shifts that translate into errors in estimated distances.

Other sources of uncertainty include the following.

1. In the application we have in mind of surveying an unknown region, possibly under adverse conditions, the placement of our sensors cannot be assumed known with such accuracy that all location errors cannot be ignored.
2. There are also uncertainties stemming from our initial ignorance of the geometry of the region we are assessing. As needed below, we will make additional assumptions concerning the appropriate distribution of microphones and that the region is a convex polyhedron with a limited number of faces.

While this dissertation takes into account some of these sources of noise, it does not account for the last-enumerated group of uncertainties.

1.3 Data Used in Reconstructing $\mathcal{R}$

The data that will be used in reconstructing $\mathcal{R}$ is a set of records of duration $[0, T]$ from each microphone that show the transient responses to a pulse emitted by the source S. At each microphone we record the initial line-of-sight pulse followed by the first echo pulse received at that microphone. Let d_1 denote the line-of-sight distance between a given microphone and the sound source and d_2 denote the distance from the given microphone to its closest image point.

$$\text{line-of-sight received pulse} \qquad \frac{A}{d_1} s\left(t - \frac{d_1}{c}\right);$$

$$\text{first echo pulse} \qquad \frac{A}{d_2} s\left(t - \frac{d_2}{c}\right).$$

As noted before, the pulse $s(t)$ is zero outside of the time interval $[0, T^*]$. Hence, the direct pulse is fully observed over a time interval of $\left[\frac{d_1}{c}, \frac{d_1}{c} + T^*\right]$. As $d_2 > d_1$, the time to fully recover the echo as well as the direct pulse requires observation over the interval $\left[0, \frac{d_2}{c} + T^*\right]$ and that this is a subset of $[0, T]$.

There are two issues. The first is that we need the direct and echo signals not to overlap—if they overlap we will not have an accurate means of determining the onset of the echo through d_2. The second issue is how long we can wait for an echo to arrive. If we have to wait too long, then the arriving signal will be so attenuated that it cannot be detected reliably by the microphone. If T is too small, then some microphones will only receive a truncated echo or none at all.

Our approach to the first issue is to select a pulse width T^* that is small enough that at no microphone do we observe overlaps between received direct and echo pulses. This may need to be done iteratively so as to select T^* small enough to ensure no overlap but large enough that we do not have to generate very intense spikes at the source to provide enough energy for accurate signal

detection.

As we do not know the maximum distance d_2 over the array of microphones, our approach to the second issue is to adopt a time upper bound of T corresponding to some large distance cT. Choice of T limits our surveillance boundaries and it may need to do so when attempts to increase T correspond to echoes that have traveled so far that the signal strength at the microphone is too weak to be detected reliably.

CHAPTER 2

DISTANCE EXTRACTION FROM WAVE SAMPLES

The first step we take to recover the room shape is to extract distance information from the recording at each microphone. As described in Section 1.3, the data acquired at each microphone consists of the line-of-sight pulse and the first echo pulse. Following the same notation as in the previous chapter, d_2 is the distance between the microphone and its closest image, and it is encoded in the first echo pulse. If we can successfully extract the delays to the first echoes from the data collected by a sufficient number of microphones, we will be able to obtain the image point locations needed to determine the room shape. In this chapter, we focus on the estimation of d_2 from a single microphone recording corrupted by noise. The details on how we use such information to reconstruct the room are provided in later chapters of this dissertation.

2.1 Introduction

2.1.1 Related Work

We notice that the extraction of d_2 requires us to detect the time-of-arrival of the first echo at each microphone. Target distance estimation has long been treated in the areas of sonar and radar applications. Our problem is like sonar, in that we have to deal with multiple echoes generated by a single sound pulse, and like radar in that we have a non-dispersive channel and our interest is in range estimation. The statistical approach usually taken in radar assumes that

the noise involved is a wideband Gaussian random process and the signal-to-noise ratio is not too small (see Skolnik [15], Helstrom [8]). In the application we have in mind, however, such assumptions are not appropriate. The noise, $n(t)$, described in Section 1.2.5, has multiple sources that do not share a common model such as a Gaussian random process. As will be clear from Subsection 2.2.2, our approach is nonparametric and, as shown in Theorem 1 of Subsection 2.2.4, it is statistically consistent in the sense that increasing the norm of the signal will decrease an upper bound on the norm of the time delay estimation error. As should be the case, this result encourages us to enlarge the source signal norm so as to achieve reasonable signal-to-noise ratio. Finally, this result shows us that we cannot tolerate excessively long waiting times to the first echo that correspond to great distances to the image points—our surveillance region is limited by our signal power, a familiar conclusion. The work in this chapter and elements of Chapter 4 are closer to traditional radar or sonar problems.

The problem of inferring a convex polyhedral room from the distance measurements, on the other hand, differs substantially from those studied in radar and sonar applications, and it is the central contribution of this dissertation. This problem is treated in Chapters 3, 5, 6 for both the noiseless and the noisy cases.

2.1.2 Problem Formulation

For clarity, let $r(t)$ denote the data from a single microphone that we use to reconstruct the room. As described in Section 1.3, the recorded signal, $r(t)$, includes the line-of-sight pulse and first echo pulse. If the additive noise $n(t)$ is as

described in Section 1.2.5, we can write:

$$r(t) = \frac{A}{d_1} s\left(t - \frac{d_1}{c}\right) + \frac{A}{d_2} s\left(t - \frac{d_2}{c}\right) + n(t),$$

where $s(t)$ is the pulse signal generated by the sound source, A is a constant representing signal amplitude that takes into account gains of the source S and the microphone and that can be adjusted at the sound source, d_1 is the line-of-sight distance between the microphone and the sound source, d_2 is the distance between the microphone and its closest image point, thereby being the path length traveled by the first echo, and c is the speed of sound in the medium. Our goal is to estimate d_2 based on the knowledge of $s(t)$, A, d_1, c, and n inferred from r.

2.2 Extracting Distances from Wave Samples

2.2.1 Assumptions

In order to solve the estimation problem, we make the following assumptions based on the models introduced in Chapter 1.

1. The acoustic pulse generated by the sound source inside the room is of duration no longer than T^* and can be modeled by a nonzero continuous function $s(t)$ that is supported on $[0, T^*]$, i.e. $s(t) = 0$ for $t \notin [0, T^*]$.
2. The total received signal $r(t)$ has support $[0, T]$, an interval which is wide enough to contain the first echo signal (see Section 1.3), i.e.

$$cT \geq d_2 + cT^*. \tag{2.1}$$

3. The delay d_1/c, to the receipt of the direct signal by the microphone, is known to us because d_1 can be measured when the system is deployed.

4. We ignore echos occurring after the first one, assuming that any of them have amplitude high enough to be detected after they have traveled a longer path from a more distant image point.

5. T^* is chosen, perhaps after initial experimentation with the deployed system, so that the direct wave and the echo signals do not overlap at any of the microphones (see Section 1.3). Nor is there temporal overlap between the first echo and a second echo, if any. Thus for a particular microphone we require that

$$d_2 - d_1 > cT^*. \tag{2.2}$$

2.2.2 Inferring the Distance to an Image Point

We ignore the direct wave and only consider

$$r_2(t) = r(t)1_{[\frac{d_1}{c}+T^*,T]}(t). \tag{2.3}$$

From Eqn. 2.2,

$$r_2(t) = \frac{A}{d_2}s(t - \frac{d_2}{c}) + n_2(t) \tag{2.4}$$

where $n_2(t) = n(t)1_{[\frac{d_1}{c}+T^*,T]}(t)$.

Define the objective function

$$L(\delta) = \|r_2(t) - \frac{A}{\delta + d_1}s(t - \frac{d_1 + \delta}{c})\| = \|\frac{A}{d_2}s(t - \frac{d_2}{c}) + n_2(t) - \frac{A}{\delta + d_1}s(t - \frac{d_1 + \delta}{c})\|, \tag{2.5}$$

where $\|\cdot\|$ is the L^2 function norm and it is evaluated with respect to t over the interval $\mathcal{D} = [\frac{d_1}{c} + T^*, T]$,

$$\|g\|^2 = \int_{\mathcal{D}} g^2(t)\,dt.$$

The lower limit is due to the fact that $r_2(t)$ is truncated to the interval $[\frac{d_1}{c} + T^*, T]$ as in Eqn 2.3. If we only record up to T in time, then δ is only meaningful when $T^* + \frac{d_1}{c} \leq \frac{d_1+\delta}{c} \leq T - T^*$; otherwise, we would have estimated a time-of-arrival earlier than the direct pulse termination or later than the observation time T. This yields the domain Δ of δ as

$$\Delta = [cT^*, c(T - T^*) - d_1]. \tag{2.6}$$

Our norm-minimizing estimate of $d_2 - d_1$ is given

$$\hat{\delta} = \arg\min_{\delta \in \Delta} L(\delta) \tag{2.7}$$

which can be calculated from our knowledge of $r_2(t)$, s and c. In the absence of noise, the zero minimum of L is achieved by $\delta^* = d_2 - d_1$ as desired. The uniqueness of this minimizing solution in the noiseless case is established next.

2.2.3 Uniqueness of Solution Without Noise

Let us assume, contrary to our expectations, that when there is no noise the minimum of L is also achieved by $\delta = d_2 - d_1 + \xi$ for some $\xi \neq 0$. Therefore,

$$L(d_2 - d_1 + \xi) = \left\| \frac{A}{d_2 + \xi} s\left(t - \frac{d_2 + \xi}{c}\right) - \frac{A}{d_2} s(t - \frac{d_2}{c}) \right\| = 0.$$

A continuous function s over a domain $\mathcal{D}$ that has a zero L^2-norm must be zero identically over $\mathcal{D}$. Hence, if there is another minimizing choice $d_2 - d_1 + \xi$ for the norm, then it must follow that

$$s(t - \frac{d_2}{c}) = \frac{d_2}{d_2 + \xi} s(t - \frac{d_2 + \xi}{c}), \quad \text{for all } \; t \in \mathcal{D} = [\frac{d_1}{c} + T^*, T], \text{ or}$$

$$s(u) = \frac{d_2}{d_2 + \xi} s(u - \frac{\xi}{c}), \quad \text{for all } \; u \in \mathcal{D}' = [\frac{d_1}{c} + T^* - \frac{d_2}{c}, T - \frac{d_2}{c}].$$

Observe that $[0, T^*] \subseteq \mathcal{D}'$ follows from the fact that T^* is chosen small enough that there is no overlap between the direct and first echo signals, implying that $d_2/c > T^* + d_1/c$, and that T is chosen large enough to capture the full echo, implying that $T > T^* + d_2/c$.

Let $\mathcal{N}$ be a subset of $[0, T^*]$ over which the function s is strictly non-zero. If $\mathcal{N}$ is empty, then s is identically zero, contradicting our assumptions. Since $\mathcal{N} \subseteq [0, T^*] \subseteq \mathcal{D}'$, we have

$$s(u) = \frac{d_2}{d_2 + \xi} s(u - \frac{\xi}{c}), \quad \text{for all } u \in \mathcal{N}.$$

Assume that $\xi > 0$. The case of $\xi = 0$ is trivial and the case of $\xi < 0$ can be treated similarly to the case of $\xi > 0$. Since $\mathcal{N}$ is a bounded non-empty set, it has a finite nonnegative infimum u_*. If u_* is in $\mathcal{N}$ then $s(u_*) \neq 0$. However, $u^* - \xi/c < 0$ and therefore $s(u_* - \xi/c) = 0$. The desired contradiction is reached. Now assume that $u_* \notin \mathcal{N}$. In this case

$$(\forall \epsilon > 0)(\exists v \in \mathcal{N}) u_* < v < u_* + \epsilon \text{ and } s(v) \neq 0.$$

However, if we choose $\epsilon < \xi/c$ then $(v - \xi/c) < 0$, not in $\mathcal{N}$, and $s(v - \xi/c) = 0$. Again, the desired contradiction is reached. For the remaining case of $\xi < 0$, the same contradiction can be reached by using the supremum u^* of $\mathcal{N}$. Hence in the noiseless case, the minimizing argument of L is unique. However, this uniqueness need not hold in the noisy case.

2.2.4 Convergence of $\hat{\delta}$ with Increasing Signal Amplitude

For given noise $n(t)$, if we continually increase the amplitude A of the source signal then we would expect the accuracy of the estimation of the delay $\hat{\delta}$ should

improve without limit. Equivalently, it is as if the noise $n(t)$ becomes continually small and the problem converges to a noiseless case.

Theorem 1 (Condition for Convergence of $\hat{\delta}$ to δ^*)**.** *The source signal $s(t)$ is as defined previously and $\delta^* = d_2 - d_1$. Assume that $\|n_2\| < \infty$. Further assume the following:*

$$(\exists C > 0, \alpha \geq 1)(\forall \delta \in \Delta) \tag{2.8}$$

$$\|s(t - \frac{d_2}{c}) - \frac{d_2}{\delta + d_1} s(t - \frac{d_1 + \delta}{c})\| \geq C|\delta - \delta^*|^\alpha. \tag{2.9}$$

It then follows that

$$\lim_{A \to \infty} \hat{\delta} = \delta^* = d_2 - d_1. \tag{2.10}$$

Prof. M. J. Todd, ORIE, Cornell University, suggested the utility of this hypothesis to achieve the desired conclusion.

Proof. For convenience, define

$$f(\delta, t) = s(t - \frac{d_2}{c}) - \frac{d_2}{\delta + d_1} s(t - \frac{d_1 + \delta}{c}), \tag{2.11}$$

for $t \in \mathcal{D} = [\frac{d_1}{c} + T^*, T]$ and $\delta \in \Delta = [cT^*, c(T - T^*) - d_1]$.

From Eqns 2.5,2.11,

$$L(\delta) = \|\frac{A}{d_2} f(\delta, t) + n_2(t)\|. \tag{2.12}$$

By the triangle inequality we have

$$\|\frac{A}{d_2} f(\delta, t)\| = \|\frac{A}{d_2} f(\delta, t) + n_2(t) - n_2(t)\| \leq \|\frac{A}{d_2} f(\delta, t) + n_2(t)\| + \|n_2(t)\|, \tag{2.13}$$

$$\|\frac{A}{d_2} f(\delta, t)\| - \|n_2(t)\| \leq \|\frac{A}{d_2} f(\delta, t) + n_2(t)\|. \tag{2.14}$$

This holds for all $\delta \in \Delta$, therefore it is true for $\hat{\delta}$,

$$\|\frac{A}{d_2} f(\hat{\delta}, t)\| - \|n_2(t)\| \leq \|\frac{A}{d_2} f(\hat{\delta}, t) + n_2(t)\|. \tag{2.15}$$

Since $\hat{\delta}$ is a global minimizer for $L(\delta)$, it is true that

$$L(\hat{\delta}) = \|\frac{A}{d_2} f(\hat{\delta}, t) + n_2(t)\| \leq \|\frac{A}{d_2} f(\delta^*, t) + n_2(t)\| = L(\delta^*). \tag{2.16}$$

Since $\|\frac{A}{d_2} f(\delta^*, t)\| = 0$ from the previous section, we have

$$\|\frac{A}{d_2} f(\hat{\delta}, t)\| - \|n_2(t)\| \leq \|\frac{A}{d_2} f(\hat{\delta}, t) + n_2(t)\| \leq \|\frac{A}{d_2} f(\delta^*, t) + n_2(t)\| = \|n_2(t)\|. \tag{2.17}$$

Consequently

$$\|\frac{A}{d_2} f(\hat{\delta}, t)\| \leq 2\|n_2(t)\|. \tag{2.18}$$

Invoking the hypothesis of Theorem 1, we conclude that

$$C\frac{A}{d_2}|\hat{\delta} - \delta^*|^\alpha \leq 2\|n_2(t)\|. \tag{2.19}$$

As a result, given that C, α, d_2 are positive constants, then for any given noise of finite norm, we have $\hat{\delta} \to \delta^*$ as $A \to \infty$. □

Hence, under the hypothesis of Theorem 1 we have shown that increasing the source signal amplitude improves the accuracy of our estimator $\hat{\delta}$ and in the limit of A diverging to infinity, the estimation error converges to zero.

2.2.5 Source Signal with Positive L^2-norm

We now turn to explore the applicability of the hypothesis of Theorem 1. Assume that the L^2-norm $\|s\|$ of the source signal $s(t)$ is positive and finite, the physically plausible case. We first prove the case for $\alpha = 1$ and then show that there is an extension to the $\alpha > 1$ case that provides a weaker lower bound. We claim that

$$\|f(\delta, t)\| \geq C_1|\delta - \delta^*|, \tag{2.20}$$

where

$$(\forall \delta \in \Delta)(\forall t \in \mathcal{D})\ f(\delta, t) = s(t - \frac{d_2}{c}) - \frac{d_2}{\delta + d_1} s(t - \frac{d_1 + \delta}{c}). \tag{2.21}$$

Once we have proven Eqn. 2.20, we will then prove an even lower bound $\|f(\delta, t)\| \geq C_\alpha |\delta - \delta^*|^\alpha$ for $\alpha > 1$.

Observe that the L^2-norm of the source signal $s(t)$ over any time set $[a, b]$, with $a \leq 0, b \geq T^*$, can be evaluated as

$$\|s\|^2 = \int_0^{T^*} s^2(u) du. \tag{2.22}$$

Recall that we have previously assumed in Section 2.2.1 that the microphone recording period $[\frac{d_1}{c} + T^*, T]$ is such that only the echo is contained in the measurement. Hence,

$$\int_{\frac{d_1}{c} + T^* - \frac{d_2}{c}}^{T - \frac{d_2}{c}} s^2(u) du = \int_0^{T*} s^2(u) du = \|s\|^2. \tag{2.23}$$

Similarly,

$$\int_{\frac{d_1}{c} + T^* - \frac{d_1 + \delta}{c}}^{T - \frac{d_1 + \delta}{c}} s^2(u) du = \int_0^{T*} s^2(u) du = \|s\|^2. \tag{2.24}$$

Now with Eqn 2.24, we can write $\|f(\delta, t)\|^2$ as the following:

$$\|f(\delta, t)\|^2 = \int_{\frac{d_1}{c} + T^*}^{T} \left(s(t - \frac{d_2}{c}) - \frac{d_2}{\delta + d_1} s(t - \frac{d_1 + \delta}{c}) \right)^2 dt \tag{2.25}$$

$$= \|s\|^2 + \left(\frac{d_2}{\delta + d_1} \right)^2 \|s\|^2 - \frac{2 d_2}{\delta + d_1} \int_{\frac{d_1}{c} + T^*}^{T} s(t - \frac{d_2}{c}) s(t - \frac{d_1 + \delta}{c}) dt. \tag{2.26}$$

By Holder's inequality we have

$$\int_{\frac{d_1}{c} + T^*}^{T} s(t - \frac{d_2}{c}) s(t - \frac{d_1 + \delta}{c}) dt \tag{2.27}$$

$$\leq \sqrt{\int_{\frac{d_1}{c} + T^*}^{T} s^2(t - \frac{d_2}{c}) dt} \sqrt{\int_{\frac{d_1}{c} + T^*}^{T} s^2(t - \frac{d_1 + \delta}{c}) dt} \tag{2.28}$$

$$= \|s\|^2. \tag{2.29}$$

Combining Eqn. 2.29 with Eqn. 2.26 yields

$$\|f(\delta,t)\|^2 \geq \|s\|^2 + \left(\frac{d_2}{\delta + d_1}\right)^2 \|s\|^2 - \frac{2d_2}{\delta + d_1}\|s\|^2 \tag{2.30}$$

$$= \left(1 - \frac{d_2}{\delta + d_1}\right)^2 \|s\|^2. \tag{2.31}$$

This yields

$$\|f(\delta,t)\| \geq \frac{|\delta + d_1 - d_2|}{\delta + d_1}\|s\|. \tag{2.32}$$

Since $\delta \in \Delta = [cT^*, c(T - T^*) - d_1]$, we have $\delta + d_1 \leq c(T - T^*)$ and we have the desired result

$$\|f(\delta,t)\| \geq \frac{|\delta + d_1 - d_2|}{c(T - T^*)}\|s\| = \frac{\|s\|}{c(T - T^*)}|\delta - \delta^*|. \tag{2.33}$$

Now assume $\alpha > 1$ and define

$$D_{\alpha-1} = \max_{\delta \in \Delta} |\delta - \delta^*|^{\alpha-1} < \infty. \tag{2.34}$$

Clearly,

$$(\forall \delta \in \Delta)\ \frac{|\delta - \delta^*|^{\alpha-1}}{D_{\alpha-1}} \leq 1.$$

Hence, from Eq. 2.33

$$\|f(\delta,t)\| \geq \frac{\|s\|}{c(T - T^*)D_{\alpha-1}}|\delta - \delta^*|^{\alpha}. \tag{2.35}$$

This result justifies the assumption made in Theorem 1 for signals of finite L^2-norm.

2.3 Examples and Simulations

2.3.1 Example One

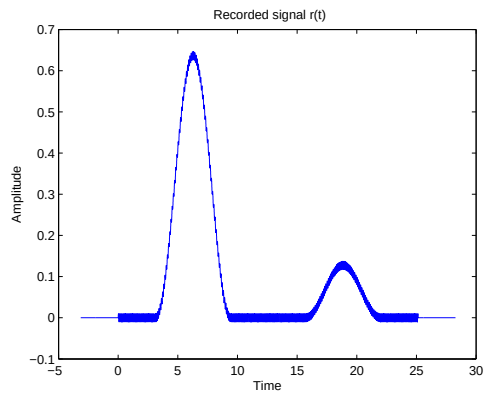
For this example, let us give the following initial values:

- Let the probe signal be $As(t)$ ($T^* = 2\pi$), where:

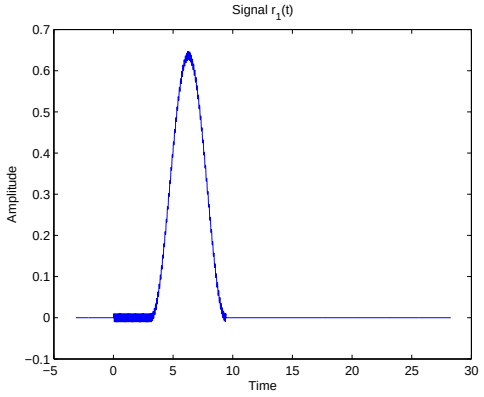
$$s(t) = \begin{cases} 1 + \cos(t - \pi), & 0 < t < 2\pi \\ 0, \text{ otherwise} \end{cases}.$$

- Let $d_1 = \pi$, $d_2 = 5\pi$ (d_2 is unknown), noise $n(t) = 0.01\sin(20\pi t)$, and duration of recording $T = 8\pi$.

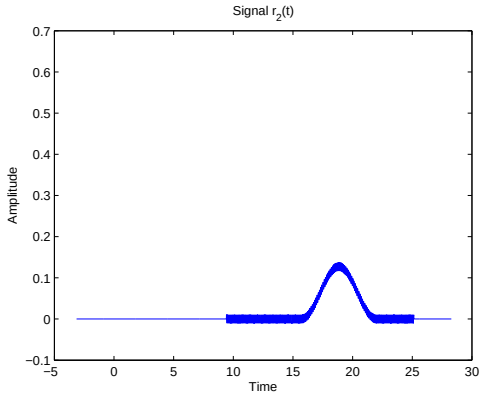
Then it follows that $\|s(t)\| = \sqrt{2\pi}$, $\|s'(t)\| = \sqrt{\pi}$, and $\|n(t)\| = 0.0354$. For simplicity, let $c = 1$, and $A = 1$. Given the probe signal and other parameters, the recorded signal $r(t)$ is as shown in Fig. 2.1(a), and signals $r_1(t)$ and $r_2(t)$ are in Fig. 2.1(b) and Fig. 2.1(c), respectively. Then $L(\delta)$ is minimal at $\hat{\delta} = 12.5$ as in Fig. 2.1(d). The estimate of d_2 is 15.6416. The error is 0.0664.



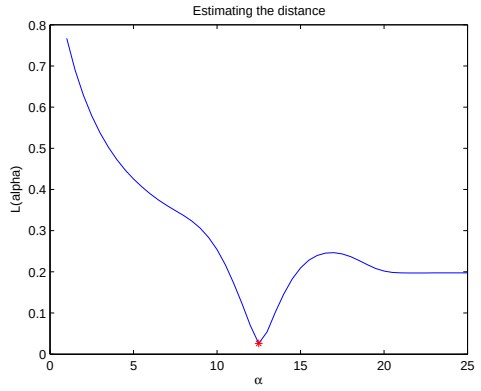
(a) Recorded signal $r(t)$



(b) $r_1(t)$



(c) $r_2(t)$



(d) $L(\delta)$

2.3.2 Example Two

For this example, let us give the following initial values, which are the same as the previous example except the noise:

- Let the probe signal be $As(t)$ ($T^* = 2\pi$), where:

$$s(t) = \begin{cases} 1 + \cos(t - \pi), & 0 < t < 2\pi \\ 0, \text{ otherwise} \end{cases}.$$

- Let $d_1 = \pi$ and $d_2 = 5\pi$ (d_2 is unknown).
- Let the noise

$$n(t) = -\frac{A}{d_1} s(t - \frac{d_1}{c}) - \frac{A}{d_2} s(t - \frac{d_2}{c}),$$

  which will result in a recorded signal $r(t) = 0$ in the entire domain.

- Let the duration of recording $T = 8\pi$.

Then it follows that $\|s(t)\| = \sqrt{2\pi}$, $\|s'(t)\| = \sqrt{\pi}$, and $\|n(t)\| = \frac{6}{5\pi}\sqrt{2\pi}$. For simplicity, let $c = 1$, and for $A = 1$. In this case, the received signal is constantly zero. Thus, applying the estimation algorithm directly would yield an infinite number of solutions, that is, any δ could be a solution. Therefore, choosing any δ in the range will result in an error that is no more than $\frac{1}{2}(c(T - T^*) - d_1)$, that is, 7.8540. This example shows a bad noise that the proposed algorithm will not work on. In practice, we may want to access the noise source first and maximize the amplitude of probe signal to achieve the best result.

2.4 Conclusion

In this chapter, we presented and defended a norm-minimizing algorithm to estimate the distance between the microphone location and its nearest source image point. Our results were nonparametric in that we only assumed a finite norm for the noise. The theoretical results were illustrated by two MATLAB simulations of the algorithm.

CHAPTER 3

NOISE-FREE IDENTIFICATION OF AN UNKNOWN ROOM

Before dealing with the noisy measurements to identify an unknown room, we start with a noiseless case where all the measurements are assumed to be noise-free. In this case, the data described in Section 1.3 contributes the exact distance measurement between each microphone and its closest image. Based on those errorless distance estimates, we are going to do the following in this chapter:

1. Provide problem setups to reconstruct a convex room from errorless distance estimates.
2. Detail the procedure of room shape identification without noise.
3. Analyze the probability of successfully recovering the room shape.

3.1 Problem Setup

To fully define the problem in a more mathematical form, we are going to provide problem models and setup in addition to the models given in Chapter 1.

3.1.1 Problem Model

This dissertation explores the use of a network of sensors to map a surveillance region, which is an unknown interior space we refer to as a room. This room is assumed to be a convex polyhedron having an unknown number W of walls or faces that are in unknown location. Our goal is to infer the geometry of this

room. Our means of inference is to measure the time of arrival of reflected signals by the walls of the room. These reflections follow the standard laws of reflection in which angles of incidence and angles of reflection are the same. We allow for partial absorption of signals by the reflective walls, although the presence of absorptive walls will decrease the size of the surveillance region. The physical nature of the signals can be either brief electromagnetic pulses or brief acoustic pulses. For economy of expression, we will refer to the acoustic version of this problem throughout. Signal amplitude is assumed to decay inversely as the distance between the emitting source and the receiving sensor.

3.1.2 Problem Setup

Throughout this chapter, we denote locations in $\mathbb{R}^3$ by boldface vectors. We have a single sound source that emits a short acoustic signal $s(t)$, $t \in [0, T^*]$, with this emission being approximately omnidirectional. This source is located at $\mathbf{0}$, the origin of our coordinate system, and its physical extent is negligible. We have a collection of $\mathcal{M}$ sensors that are nearly omnidirectional microphones. The sensors and source are connected by a network to a processing point, with delays in this network being negligible compared with T^*. The sensors occupy known positions, with sensor i at $\mathbf{d}_i$ and $d_i = |\mathbf{d}_i|$. The distance between sensors i and j is $d_{i,j} = |\mathbf{d}_i - \mathbf{d}_j|$. As the distance d_i from the source to sensor i is known, and we also know the speed of sound c, we can accurately predict the time of arrival of the direct signal from the source. The measurement made by a sensor is to record its environment over a total time period of duration T. The i-th sensor searches, for an echo of the known transmitted pulse s of duration T^*, over a time interval $[T^* + d_i/c, T]$ that spans from the known time of receipt

of the direct signal to a chosen upper bound T. cT determines the maximum distance travelled by the first echo from emission at time 0 at location $\mathbf{0}$ to its arrival at sensor i. We have effectively limited our surveillance region roughly to a sphere of radius cT. It is possible that multiple echoes might be received at certain sensors within the observation period T. In our data analysis we ignore all but the first echo. This record at each sensor is then processed to extract the time of arrival of the acoustic signal from the source and the time of arrival of the first echo of this source signal. These times of arrivals are converted into distances ρ_i corresponding to the path lengths from the source to the sensors via reflection off the nearest walls of the room.

3.2 Room Identification

3.2.1 Wall and Room

Assume throughout that the point-like sound source is located at the origin of the coordinate system. An *image point* $\mathbf{I}$ determines a plane π specified by a normal vector $\mathbf{I}$ and a point $\frac{1}{2}\mathbf{I} \in \pi$ through:

$$\pi = \{\mathbf{x} : \mathbf{I} \cdot \mathbf{x} = \frac{1}{2}||\mathbf{I}||^2\},$$

where $||\mathbf{x}||$ denotes the Euclidean length of $\mathbf{x}$. A *wall* is a convex polygonal subset of such a plane π.

Assuming that there are W walls, with W initially unknown to us, then these W walls are polygonal subsets of $\{\pi_i : i = 1 : W\}$ determined by $\{\mathbf{I}_i, i = 1 : W\}$, which are initially also unknown to us. Let H_i denote the half-space containing

the sound source at $\mathbf{0}$ that is determined by the plane π_i through:

$$H_i = \{\mathbf{x} : \mathbf{I}_i \cdot \mathbf{x} \leq \frac{1}{2}\|\mathbf{I}_i\|^2\}.$$

The convex room is

$$\mathcal{R} = \bigcap_{i=1}^{W} H_i.$$

Failure to identify any of the image points, possibly because there are too few microphones, will mean that the estimated room is a convex polyhedron that is larger than it should be due to the absence of one or more limiting half-spaces from the intersection defining the room.

3.2.2 Image Point and Voronoi Regions

To each wall w_k of the room, we correspond an image point $\mathbf{I}_k$ of the acoustic source at $\mathbf{0}$. This image point is such that a reflection (echo) off wall w_k can be treated as if the acoustic source was physically located at $\mathbf{I}_k$ and there was straight-line propagation from $\mathbf{I}_k$ to each sensor. The first echo received at sensor i can be treated as having come from the image point that is closest to $\mathbf{d}_i$. There are as many image points $\{\mathbf{I}_i\}$ as there are walls in the room we are attempting to reconstruct through sonic probing. As a *first* echo received at a given sensor can only come from the image point nearest to that sensor, we are interested in the regions of points closest to any given image point.

A *Voronoi region or set* is a concept from quantization that applies to metric spaces. In our case, our metric space is $\mathbb{R}^3$ with the Euclidean metric $\|\cdot\|$.

Definition 1. *(Voronoi Region) Given a set of points/centers* $\{\mathbf{c}_i\}$*, the Voronoi region* V_i *containing just* $\mathbf{c}_i$ *is the set of points in* $\mathbb{R}^3$ *that are at least as close to* $\mathbf{c}_i$ *as to any other*

$\mathbf{c}_k$;

$$V_i = \{\mathbf{x} \in \mathbb{R}^3 : (\forall k)\ ||\mathbf{x} - \mathbf{c}_i|| \leq ||\mathbf{x} - \mathbf{c}_k||\}\ and\ \cup_i V_i = \mathbb{R}^3. \tag{3.1}$$

We have arbitarily chosen $\leq$ rather than $<$, and this choice means that our Voronoi regions overlap. A Voronoi region is a convex polyhedron. Regions of overlap are subsets of two-dimensional planes.

Given a microphone at location $\mathbf{d}$ in Voronoi region V determined by an image point or center $\mathbf{I}$, we are interested in the spherical surface

$$S(\mathbf{d}, ||\mathbf{I} - \mathbf{d}||) = \{\mathbf{x} : ||\mathbf{x} - \mathbf{d}|| = ||\mathbf{I} - \mathbf{d}||\}. \tag{3.2}$$

From our measurement of the time delay $t = ||\mathbf{d} - \mathbf{I}||/c$ to the receipt of the first echo at the microphone at $\mathbf{d}$, and our knowledge of the speed of propagation of sound c, we learn from a single microphone that

$$||\mathbf{I} - \mathbf{d}|| = ct = d_2. \tag{3.3}$$

Hence,

$$S(\mathbf{d}, d_2) = \{\mathbf{x} : ||\mathbf{x} - \mathbf{d}|| = d_2\}, \mathbf{I} \in S(\mathbf{d}, d_2), \tag{3.4}$$

is the set of points at the same distance d_2 from a center $\mathbf{d}$. We associate a set $\{S_i(\mathbf{d}_i, ||\mathbf{I} - \mathbf{d}_i||)\}$ of spherical surfaces to a collection of microphone locations $\{\mathbf{d}_i\}$ sharing a single Voronoi region with an image point $\mathbf{I}$,

$$S_i(\mathbf{d}_i, ||\mathbf{I} - \mathbf{d}_i||) = \{\mathbf{x} : ||\mathbf{x} - \mathbf{d}_i|| = ||\mathbf{I} - \mathbf{d}_i||\}.$$

We infer the image point $\mathbf{I}$ when it is the unique point of the intersection $\cap_i S_i(\mathbf{d}_i, ||\mathbf{I} - \mathbf{d}_i||)$.

3.2.3 Sensor Locations and Voronoi Regions

Sensor numbers and placements are at our disposal. We may either dispose them deterministically (e.g., at lattice points or at the vertices of Platonic/Archimedean solids) or randomly according to a given distribution. The analysis we engage in will assume that sensors are placed on a spherical surface $S(\mathbf{0}, r)$, centered on the sound source and having a radius r, in accordance with a Poisson process of mean rate λ per unit area. The expected number M and variance σ^2 of the number of such sensors are then:

$$M = 4\pi r^2 \lambda \text{ and } \sigma^2 = M \text{ with } \frac{M}{\sigma} = 2r\sqrt{\pi\lambda}.$$

Hence, for $r\sqrt{\lambda} >> 1$, the value M is, with high probability, close to the actual random number of sensors placed according to the Poisson process. If we partition $S(\mathbf{0}, \mathbf{r})$ into $\{S_i\}$ having respective areas of $\{A_i\}$, then the number of sensors in S_i is again Poisson with parameter $A_i\lambda$ and is independent of the number of sensors assigned to the other $\{S_j, j \neq i\}$.

The intersection of a half-space with a sphere results in a *spherical cap*. When the half-space completely contains the sphere, then the spherical cap is the whole sphere. Hence, the intersection of the Voronoi region V_i with $S(\mathbf{0}, r)$ is an intersection of spherical caps restricted to $S(\mathbf{0}, r)$. The intersection of two spherical caps is a lens-shaped region defined by its two extreme cusp points. Let W_r^* denote the number of image points whose Voronoi regions have a nonempty intersection with $S(\mathbf{0}, r)$. W_r^* is nondecreasing in r and cannot exceed the unknown W. Define a *spherical convex polygon* k as a connected subset of $S(\mathbf{0}, r)$ having a boundary of a finite number of circular arcs (typically, small circles and not great circles) on $S(\mathbf{0}, r)$ and satisfying the condition that all points $\mathbf{x}, \mathbf{y}$ in k are connected by an arc of a great circle that lies completely within k. $V_i \cap S(\mathbf{0}, r)$

is a spherical convex polygon k_i. If, for some i, $V_{i,j} \supseteq S(\mathbf{0}, r)$, then the image point I_j can never be observed as a first echo by any sensor positioned on $S(\mathbf{0}, r)$.

3.2.4 Learning the Room

It is possible for all of the walls to be so far from all of the sensors that no echoes are detected in time T. If that proves to be the case, then we could increase T, subject to the constraint that a long-delayed echo would have travelled a correspondingly long distance and might be sufficiently attenuated to render inaccurately the extraction of the echo transit time.

We have no access to information on how the Voronoi regions that do intersect the spherical surface partition that surface into regions $\{S_i\}$. What we do know is which of the random number of deployed sensors recorded an echo from which the distance to an image point was inferred, but not a direction to the image point. If in fact we had deployed M sensors of which M^* received echoes, then our critical task is to *cluster* these sensors according to their correspondence with the image points.

In the near absence of noise, we idealize and look for groups of at least four sensors that correspond to the same image points. The number "four," in the generic case, is the minimum number of intersecting spheres that have a unique point of intersection. Of course, we may have no observations of some image points and less than four observations of others.

Let ρ_i represent the distance from sensor i to its nearest image point. Assuming that an echo has been received, ρ_i is known from the measured time to

the arrival of a first echo. The locus of the image point producing a first echo at sensor i is the surface of the sphere centered at $\mathbf{d}_i$ and having radius ρ_i. It is obvious that either $|\rho_i - \rho_j| > d_{i,j}$ or $d_{i,j} > \rho_i + \rho_j$ assures us that these two sensors have detected different image points or walls. While this condition is sufficient for such a conclusion, it is not necessary.

When the spherical surfaces $S(\mathbf{d}_i, \rho_i), S(\mathbf{d}_j, \rho_j)$ have a nonempty intersection, then that intersection is generically a small circle. If sensors i, j are responding to the same image point, then the resulting small circle is the locus of that echo-producing image point. However, there can be a small circle intersection even when the two surfaces correspond to echoes from two *different* image sources. We simply need:

$$\rho_i + \rho_j > d_{i,j} \text{ and } \rho_i + d_{i,j} > \rho_j \text{ and } \rho_j + d_{i,j} > \rho_i$$

for a small circle intersection.

In the case of three responses at sensors i, j and k, if these responses come from a common image source, then the three spherical surfaces must generically intersect at two points. When there are four different sensor responses induced by a common image point, then the intersection of all four spherical surfaces is a single point that is the unique image point. However, even in the case of four spherical surfaces that intersect at a single point, there is no logical necessity to conclude that there is a single common image point. We can as well have four distinct image points, one on each of the four spherical surfaces that intersect in a single point, giving rise to the same picture.

3.2.5 Classifying Microphone Placements – General Position and Genericity

We need to characterize the "general case" of microphone placement in order to prove subsequent theorems about our ability to infer image points from responses of these microphones at known locations. There are two aspects to this "general case" – general position and genericity. General position is a geometric requirement that is somewhat brittle in that its classification can be changed by infinitesimal changes in locations. Genericity refers to invariance of inferences as the microphones are moved over regions of positive volume. Genericity is a more robust concept than general position but lacks the precision of general position.

General Position

Definition 2 (General Position)**.** *Given* $\mathbb{R}^3$ *containing* k *microphones at positions* $\{\mathbf{d}_i : i = k\}$ *and a single image point at* $\mathbf{I}$*, we say that they are in* general position *if the following hold:*

GP1 *If* $k = 1$*, then* $\mathbf{d}_1$ *and* $\mathbf{I}$ *are not coincident.*

GP2 *If* $k = 2$*, then* $\mathbf{d}_1, \mathbf{d}_2$ *and* $\mathbf{I}$ *are not collinear.*

GP3 *If* $k = 3$*, then* $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ *and* $\mathbf{I}$ *span three-dimensional space.*

GP4 *If* $k = 4$*, then* $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4$ *span three-dimensional space and their subsets satisfy the three preceding conditions GP1, GP2, GP3.*

GP5 *If* $k > 4$ *then all subsets microphone locations of sizes less than or equal to four satisfy the four preceding conditions.*

Genericity

The results of this and the next two sections are of interest mainly in the noiseless case. We use the terms *generic (case)* and *genericity* to refer to a condition on the microphone locations within a Voronoi region. Let μ_j denote the number of microphones located in Voronoi region V_j generated by image point $\mathbf{I}_j$. Let ν denote the number of Voronoi regions, which equals the number of image points. Our notion of a generic case will be relative to our method for inferring the image points that are the centers of the Voronoi regions.

Definition 3. *(Generic Case) The* generic case *holds for our inferences about image point locations that are based upon distances to microphones at known locations if and only if these inferences are invariant when for all* $1 \leq j \leq \nu$ *the* μ_j *microphones located in Voronoi region* V_j *are moved anywhere within a* positive volume *subset* Σ_j *of* $(\mathbb{R}^3)^{\mu_j}$*—the* $3\mu_j$*-dimensional set of location coordinates of all* μ_j *microphones lying within* V_j*.*

The restriction to a "positive volume" subset could as well be slightly generalized to a subset having positive probability with respect to a probability measure that is absolutely continuous with respect to Lebesgue measure.

As an example of a Σ_j, let $\mu_j = 2$ and $V_j = [0,1]^3$ be the unit cube and $(0,1)^3$ denote the interior of V_j. In this case we could have

$$\Sigma_j = [(0,1)^3 \times (0,1)^3]$$
$$- \{(x_1, x_2, x_3, y_1, y_2, y_3) : 0 < x_1 < 1, 0 < x_2 < 1, 0 < x_3 < 1, y_1 = x_1, y_2 = x_2, y_3 = x_3\}.$$

Σ_j uses this interior for the pair of microphone locations after first deleting the pairs of equal points, as needed to maintain distinctness between microphone locations. The volume of $\Sigma_j = 1$, as the volume is zero for the set of points at

least one of which is on the boundary of V_j and is zero for the set of pairs of points that are equal.

What Can be Inferred from Distances to Microphones Located Within a Single Voronoi Region?

As we are in a single Voronoi region, there is only one image point located at $\mathbf{I}$. In this section we will only use the condition of general position of microphones. The implications of general position are stated next, with only the next to last case needing explicit proof.

1. We cannot learn the location of $\mathbf{I}$ in the Voronoi region V from the distance from $\mathbf{I}$ to a single microphone at $\mathbf{d}_1$ unless we violate GP1 of the preceding section and have $\mathbf{d}_1 = \mathbf{I}$. We are then effectively reduced to the zero-dimensional case.
2. We cannot learn the location of $\mathbf{I}$ in the Voronoi region V from the distances from $\mathbf{I}$ to a pair of microphones $\mathbf{d}_1, \mathbf{d}_2$ in V unless we violate GP2 and have $\mathbf{d}_1, \mathbf{d}_2, \mathbf{I}$ be collinear. We are then effectively reduced to the one-dimensional case.
3. We cannot learn the location of $\mathbf{I}$ in V from distances from $\mathbf{I}$ to a trio of microphones $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ in V unless we violate GP3 and have $\mathbf{I}, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ be coplanar. We are then effectively reduced to the two-dimensional case.
4. We can learn the location of $\mathbf{I}$ in V from distances from $\mathbf{I}$ to four microphones in V when their locations satisfy GP4.
5. We can learn the location of $\mathbf{I}$ in V from distances from $\mathbf{I}$ to more than four microphones in V when their locations satisfy GP5.

The last case follows from the next to last case. Hence, only the next to last claim requires proof.

Theorem 2 (Identifying a unique image point when GP4 holds)**.** *Assume that inside a given Voronoi region we have four microphones at* $\{\mathbf{d}_i : i = 1 : 4\}$ *and an image point* $\mathbf{I}$*, and that the five points satisfy GP4. We assert that the intersection* $\cap_1^4 S_i$ *is precisely* $\mathbf{I}$*.*

Proof. It is immediate from the definition of the spherical surfaces that the point $\mathbf{x} = \mathbf{I}$ is on each surface. Hence, $\cap_1^4 S_i$ is nonempty and contains at least $\mathbf{I}$. To verify the uniqueness of this point of intersection, assume to the contrary of our belief that there is another point $\mathbf{J} \neq \mathbf{I}$ in $\cap_1^4 S_i$. Hence,

$$(\forall i = 1 : 4)\ \|\mathbf{J} - \mathbf{d}_i\| = \|\mathbf{I} - \mathbf{d}_i\|. \tag{3.5}$$

This asserts that each $\mathbf{d}_i$ is equidistant from $\mathbf{J}$ and $\mathbf{I}$ and therefore each must lie on the unique plane that perpendicularly bisects the nonempty line segment joining $\mathbf{J}$ and $\mathbf{I}$. However, by GP4, these four points cannot be coplanar. The desired contradiction is reached. □

What Can Be Inferred When Four Microphones Are Distributed Over More than One Voronoi Region?

We will need genericity to prove the following

Theorem 3 (Non-existence of a unique common point of intersection when there are four microphones distributed over more than one Voronoi region.)**.** *Assume that we have four microphones with locations and nearest image point pairs* $\{(\mathbf{d}_i, \mathbf{I}_i) : i = 1 : 4\}$*, corresponding spherical surfaces* $\{S_i : i = 1 : 4\}$*, and that these four microphones*

are located in more than one Voronoi region. When genericity holds for the microphone locations, then $\cap_1^4 S_i$ *is empty.*

Proof. We first use genericity to eliminate the need to consider the case that the same microphone location $\mathbf{d}$ might be equidistant from two image points $\mathbf{I}_1, \mathbf{I}_2$ corresponding to different Voronoi regions V_1, V_2; i.e., $\mathbf{d}$ lies on the boundaries of both of these regions. As the Voronoi regions differ, so must the image points they contain differ. If, contrary to our expectation,

$$\|\mathbf{d} - \mathbf{I}_1\| = \|\mathbf{d} - \mathbf{I}_2\|,$$

then, as the image points are distinct, $\mathbf{d}$ is constrained to lie on the plane H that is the perpendicular bisector of the non-degenerate line segment joining $\mathbf{I}_1, \mathbf{I}_2$. Hence, $\mathbf{d}$ is constrained to lie on a surface of zero volume, this is not the generic case, and, henceforth, it will be neglected.

There are several cases. We first prove the theorem for case where three microphones are in, say, V_1, and the fourth is in Voronoi region V_2. Hence, $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ are all closest to the image point $\mathbf{I}_1$ and $\mathbf{d}_4$ is closest to $\mathbf{I}_2$. If in addition, contrary to our belief, there is a point $\mathbf{J}$ common to all four spherical surfaces, then

$$\text{for } i = 1:3, \|\mathbf{d}_i - \mathbf{J}\| = \|\mathbf{d}_i - \mathbf{I}_1\| \text{ and } \|d_4 - \mathbf{J}\| = \|d_4 - \mathbf{I}_2\|. \tag{3.6}$$

If we assume that $\mathbf{J} = \mathbf{I}_1$, then

$$\|\mathbf{d}_4 - \mathbf{J}\| = \|\mathbf{d}_4 - \mathbf{I}_1\| = \|\mathbf{d}_4 - \mathbf{I}_2\|,$$

and $\mathbf{d}_4$ is same distance to $\mathbf{I}_1$ and $\mathbf{I}_2$. This restricts the placement of $\mathbf{d}_4$ to be on the overlap of Voronoi region V_1 and V_2, that is a subset of two-dimensional plane with zero volume. Such constrain is a failure of genericity. The same argument leads us to the conclusion that if $\mathbf{J}$ exists then $\mathbf{J} \neq \mathbf{I}_2$.

If $\mathbf{J}$ exists then it is distinct from $\mathbf{I}_1, \mathbf{I}_2$. Therefore, there are two planes H_1, H_2 that are the perpendicular bisectors of the nonempty line segments from $\mathbf{J}$ to $\mathbf{I}_1$ and $\mathbf{J}$ to $\mathbf{I}_2$, respectively. H_1 contains $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ and H_2 contains $\mathbf{d}_4$. We need go no further, as constraining three microphone locations to a plane H_1 is already a failure of genericity.

Let us now assume a second case in which $\mathbf{d}_1, \mathbf{d}_2$ are in V_1 with image point $\mathbf{I}_1$ and $\mathbf{d}_3, \mathbf{d}_4$ are in V_2 with image point $\mathbf{I}_2$. The argument given for the first case suffices to show that if $\mathbf{J}$ exists in $\cap_1^4 S_i$ then it cannot equal either of $\mathbf{I}_1, \mathbf{I}_2$. Nonetheless, if $\mathbf{J}$ exists, then again there are two planes H_1, H_2 with two microphone locations on each. As before, this planar constraint on the microphone locations violates genericity, and we conclude that $\cap_1^4 S_i = \emptyset$.

The third case has two microphones at $\mathbf{d}_1, \mathbf{d}_2$ located in Voronoi region V_1 and the remaining two microphones at $\mathbf{d}_3$ in Voronoi region V_3 and $\mathbf{d}_4$ in Voronoi region V_4. As the key to the proof is that when two microphones are associated with a single image point and there exists $\mathbf{J}$ unequal to either $\mathbf{I}_1$ or $\mathbf{I}_3$ or $\mathbf{I}_4$, then the microphone locations at $\mathbf{d}_1, \mathbf{d}_2$ are constrained to lie on H_1 and this alone is a failure of genericity.

The fourth and final case is that each $\mathbf{d}_i$ lies in a distinct Voronoi region V_i. By the same argument as above, we conclude if there exists $\mathbf{J}$ as the intersection of S_i, then $\mathbf{J}$ is distinct from $\mathbf{I}_i$, $i = 1:4$. Again it can be shown that $\mathbf{d}_1$ is restricted to be on the plane H_1, hence $\cap_1^4 S_i$ being nonempty violates genericity. □

We have not discussed the more general case where we have more than four microphones, yet less than four in any one Voronoi region. Given the genericity condition, it is likely, from the arguments given in the proof of the immedi-

ately preceding theorem, that the intersection of at least four spheres $\{S_i\}$ will be empty.

3.2.6 Identifying a Room from Sensor Responses

Now we are ready to give the algorithm to identify a room from noiseless sensor responses.

Setup

1. There is a set of sensors $\mathcal{M} = \{s_i\}$, $\|\mathcal{M}\| = M$, located at $\{\mathbf{d_i}\}$.
2. If sensor s_i detects an echo from a closest image point I, then the time delay enables inference to the distance ρ_i from s_i to that image point. In the noiseless case of Chapter 2, this identification is exact.
3. If the closest image point to sensor s_j is so distant, that is, approximately at least T/c, then no echo is detected in the observation interval ending at T. The absence of a report from s_j cannot contribute to locating such possible image points beyond noting their minimal distance, if they exist. Let $\mathcal{M}^* \subseteq \mathcal{M}$ denote the set of sensors that have received an echo signal.
4. Henceforth, we restrict our attention to the sensors in $\mathcal{M}^*$.
5. $S(\mathbf{d_i}, \rho_i)$ denotes the spherical surface in a three-dimensional space with its center at $\mathbf{d_i}$ and radius ρ_i such that it contains a unique image point. While it is mathematically possible for there to be more than one image point at the same distance from $\mathbf{d_i}$, this case is not generic. Subsequently, when we consider disposing sensors according to a certain random distribution,

the conditional probability will be zero for more than one of the finite set of image points being at the same distance from any $\mathbf{d_i}$, given that there already is an image point at that distance.

Identifying Image Points of Unknown Location

The task is to identify the image points (and therefore the walls of the room) and to partition the set $\mathcal{M}^*$ of sensors according to their nearest neighbor image points. In some instances, this partitioning or clustering will yield unique identifications of the image points, but not in all cases. There will be unique identification if and only if at least four sensors are closer to the same image point than to any other image point.

1. Consider the $\binom{M^*}{2}$ of unordered pairs $\mathcal{P}_2 = \{[(\mathbf{d_{i_1}}, \rho_{\mathbf{i_1}}), (\mathbf{d_{i_2}}, \rho_{\mathbf{i_2}})]\}$ of distinct responsive sensors.
2. For each pair in $\mathcal{P}_2$, determine whether or not the corresponding spherical surfaces have a nonempty intersection. As shown next, this can be done easily without having to identify this intersection.
3. There are two cases that cover the possibilities of having no intersection between two spherical surfaces:
   (a) (Condition 1) Disjoint corresponding spheres: $\rho_{i_1} + \rho_{i_2} < \delta$.
   (b) (Condition 2) One sphere contains the other: $|\rho_{i_1} - \rho_{i_2}| > \delta$.
4. Failure of both conditions is the necessary and sufficient condition for there to be an intersection between the spherical surfaces. Thus, failure is easily determined and simpler than solving for the actual (generally small) circle of intersection.

5. Let Q_2 be the subset of $\mathcal{P}_2$ of pairs of spherical surfaces that have a nonempty intersection. Calculation of Q_2 requires $\binom{M^*}{2}$ examinations of the failure of Conditions 1 and 2.

6. From Q_2, determine the set Q_3 of triples of spherical surfaces such that each pair in the triple has a nonempty intersection. The size of Q_3 is no greater than $\binom{M^*}{3}$.

7. The computational effort required to identify Q_3 is no greater than proportional to $\binom{M^*}{3}$. For each triple of sensors, we take a look at each of the three pairs in that triple to see if they are in Q_2.

8. For each triple of spherical surfaces in Q_3, solve for the (at most) two generic points of intersection. Some triples will have no common points of intersection. The others will generically have two points. The two non-generic cases are:

   (a) A single point of intersection when a circle determined by two sensors has a single point of intersection with a spherical surface determined by a third sensor;

   (b) Infinitely many points of intersection when the circle determined by two sensors is embedded in the spherical surface determined by a third sensor.

9. Let Q_3^* denote the subset of Q_3 consisting of triples of those spherical surfaces whose intersection contains exactly two points.

10. At least one of each pair of points of intersection found for a triple of sensors is a candidate for a sensor location.

11. Let $\mathcal{P}_5$ be the set of 5-tuples consisting of a triple of sensor locations from Q_3 followed by their (generically) two associated points of the common

intersection of their three corresponding spherical surfaces.

12. For each quintuple in $\mathcal{P}_5$, and for each sensor response $(\mathbf{d_k}, \rho_\mathbf{k})$ not included in that quintuple, determine if either of the two points of intersection in the quintuple lies on the spherical surface $S(\mathbf{d_k}, \rho_\mathbf{k})$. If one of these points does lie on this surface, then it is an image point, and it is the closest image point to these four sensors. Form the list Q_5 of quintuples consisting of a quadruple of sensor locations followed by the identified image point common to all four sensor responses.
13. Q_5 immediately yields the desired set $\mathcal{I}^*$ of identified or located image points.

Estimating the Number of Image Points of Unknown Locations

Once we obtain the desired set $\mathcal{I}^*$, what can we say about the image points that are known to exist but are not in $\mathcal{I}^*$?

1. Form a set $\mathcal{U} \subset \mathcal{M}^*$ of those remaining sensors whose associated spherical surfaces do not contain any of the image points in $\mathcal{I}^*$. No sensor in $\mathcal{U}$ has a spherical surface that generates a unique point of intersection with any other spherical surfaces generated by the sensors in the quintuples of Q_5.
2. All of the sensors in $\mathcal{U}$ have responded to an echo from an unidentified image point, one that is not in $\mathcal{I}^*$.
3. If sensor $s_k \in \mathcal{U}$, then there is an unidentified image point that is precisely at a distance ρ_k from $\mathbf{d_k}$.
4. $\mathcal{U}$ provides some information about the number of undiscovered image points.

5. If sensor $s_k \in \mathcal{U}$ has a spherical surface $S(\mathbf{d_k}, \rho_\mathbf{k})$ that is disjoint from the spherical surfaces corresponding to any of the other sensors in $\mathcal{U}$, then s_k corresponds to an image point lying on $S(\mathbf{d_k}, \rho_\mathbf{k})$ that is distinct from the image points lying on the spherical surfaces of any other sensor in $\mathcal{U}$ (or in $\mathcal{I}^*$).

6. If there is an intersection of the two spherical surfaces of two sensors in $\mathcal{U}$, then we do not know whether that intersection contains a location of a single image point closest to both of those sensors or whether this arises from two distinct image points.

7. If there is an intersection of the three spherical surfaces of three sensors in $\mathcal{U}$, then we do not know that the two points of this intersection correspond to image points.

3.3 Probability Modeling for Room Identification

The procedure introduced in Section 3.3 provides a direct means to recover a room based on noise-free measurement of distances. However, as the room is unknown at the outset, it is neither guaranteed that we have four microphones placed in each Voronoi region nor that every Voronoi region is detected. Therefore, it is crucial to investigate what is the probability to recover a room.

3.3.1 Models for the Sensor Set M

We can model the sensor set M as deployed either deterministically or probabilistically.

A Deterministic Structure for the Sensor Set M

1. We assume that all elements of M are also extreme points of the convex hull of M, although this is not necessary from a mathematical viewpoint. We assert this to move the sensors closer to the image points they are to detect so as to receive stronger echo signals.

2. Given the absence of any knowledge as to the locations of the image points, there are no preferred directions (as noted in Section 1.2.3). For purposes of analysis, we place the constellation M on the surface $S(\mathbf{0}, \underline{r})$ of a sphere of radius $\underline{r}$ centered at the sound source at $\mathbf{0}$.

3. The symmetry of our prior information suggests a design in which all elements of M are at the same distance from their nearest neighbors. There are no "preferred" vertices in the sense that they are to be *congruent under rotation*.

4. The implication of the two preceding assumptions is that M is the set of vertices of either a *Platonic* or an *Archimedean solid* (see Cromwell [3]). Note that item 3 is true for Plantonic only.

5. There are only five Platonic solids, and they are completely regular convex polyhedra: tetrahedron, cube, octahedron, dodecahedron and icosahedron.

6. There are only 13 (15 counting left-hand, right-hand distinctions) Archimedean solids. They are convex polyhedra, and they all have the property that their vertices are congruent.

7. Hence, all vertices are at the same distance from their nearest neighbors, and the same kinds of faces meet at each vertex, in the same clockwise arrangement, when viewed from the interior of the solid.

8. Identification of an image point **I** generically requires four sensors located in M (see Section 3.2.5).

9. Restrict M to be the vertex set of either a Platonic or Archimedean solid having a good number of vertices.

10. As four is the fewest faces of a positive volume-containing convex polyhedron, and we need a minimum of four vertices to identify each face/wall, we need to choose a solid having at least 16 vertices.

11. A candidate is the *truncated cuboctahedron* having 48 vertices, 26 faces, and 72 edges, with each vertex having a (4.6.8) pattern of faces. Of course, this assumes we can deploy 48 microphones in this pattern.

Poisson Point Process Model for M

1. In the remainder of this chapter, we analyze the probability that our system will infer a convex polyhedral set $\mathcal{R}^*_{\underline{r}}$ that contains the room $\mathcal{R}$, assuming that M is a random set distributed over the spherical surface $S(\mathbf{0}, \underline{r})$.

2. More specifically, this random set M is generated by a Poisson point process on the spherical surface $S(\mathbf{0}, \underline{r})$ with per-unit area intensity parameter λ. The cardinality of M:

$$|M|, \text{ is distributed as } \mathcal{P}(\lambda A) \text{ where } A = 4\pi \underline{r}^2.$$

3. The number of sensors M_i in any subarea S_i, of size A_i, of S is also Poisson with parameter λA_i. Disjoint subareas S_i, S_j give rise to independent numbers of sensors M_i, M_j.

3.3.2 Probability Model for I

For the sensor model of the Poisson point process, we need to know the number $W^*_{\underline{r}}$ of image points whose Voronoi regions have a nonempty intersection with the spherical surface $S(\mathbf{0}, \underline{r})$ and partition $S(\mathbf{0}, \underline{r})$ into the set of regions $\{S_1, \ldots, S_{W^*_{\underline{r}}}\}$ having the set of respective areas $\{A_1, \ldots, A_{W^*_{\underline{r}}}\}$. How the Voronoi regions dispose themselves depends upon the unknown locations of the image points in $\mathcal{I}$. Our lack of prior information about the disposition of the Voronoi regions is represented by our indifference to the actual order of enumeration of these regions. While we know that

$$(\forall \underline{r}) W^*_{\underline{r}} \leq W \text{ and } \underline{r}' > \underline{r} \implies W^*_{\underline{r}'} \geq W^*_{\underline{r}},$$

we have no more precise knowledge of $W^*_{\underline{r}}$ than that. We may attempt to model the set $\mathcal{I}$ of image points as a Poisson point process in some thick spherical shell centered at the origin and bounded away from its interior spherical surface $S(\mathbf{0}, \underline{r})$ containing the sensors. If we do so, then the number W of image points will indeed be a Poisson random variable. However, we do not do this because a Poisson point process model for $\mathcal{I}$ will have a positive probability of generating a non-convex polyhedral room that cannot be determined as an intersection of half-spaces. Rather than directly modeling $\mathcal{I}$, in the next section we will make an additional modeling assumption about the intersections of the Voronoi regions generated by $\mathcal{I}$ with $S(\mathbf{0}, \underline{r})$, as this is the information we need.

3.3.3 Modeling the Voronoi Region Sets $\{S_i\}$

Our very limited prior knowledge of the disposition of the walls or, equivalently, of $\mathcal{I}$ is such that our probability model, for the subsets $\{S_i\}$ of the spheri-

cal surface $S(\mathbf{0}, \underline{r})$ corresponding to the intersections with the Voronoi regions, is a function only of their associated areas $\{A_i\}$, and furthermore, the joint probability density function of these areas is permutation-invariant or exchangeable. The joint distribution of the $\{A_i\}$, conditional upon there being $W^*_{\underline{r}} = w$ Voronoi regions intersecting $S(\mathbf{0}, \underline{r})$, is denoted by:

$$f_{A_1,\ldots,A_{W^*_{\underline{r}}}|W^*_{\underline{r}}=w}(a_1, \ldots, a_w).$$

And it is a symmetric function of its $W^*_{\underline{r}} = w$ arguments. As rotation does not change the area of a surface, this joint distribution for the $\{A_i\}$ is invariant with respect to a rotation of the spherical surface $S(\mathbf{0}, \underline{r})$ containing the microphones.

3.3.4 Dirichlet Distribution Model

To choose a specific model for $f_{A_1,\ldots,A_{W^*_{\underline{r}}}|W^*_{\underline{r}}=w}(a_1, \ldots, a_w)$ that can give us some traction on a probabilistic understanding of inferring a room, we first note that:

$$(\forall i)\ A_i \geq 0 \text{ and } \sum_{i=1}^{W^*_{\underline{r}}} A_i = A = 4\pi\underline{r}^2.$$

The summation condition forces the areas to be dependent random variables. Except for the not necessarily unit value of A, $\{A_1, \ldots, A_{W^*_{\underline{r}}}\}$ is formally analogous to a probability mass function $\{p_i = A_i/A\}$ for an experiment with $W^*_{\underline{r}}$ possible outcomes.

Choosing *priors* for such pmfs is much discussed in Bayesian statistical analysis of discrete data sources where the Bayesian statistician needs a prior probability density over the space of possible pmfs so that she can calculate a posterior probability after making observations and knowing the likelihood for observations given a pmf. The *generalized Dirichlet distribution* has a vector

$\theta = (\theta_1, \ldots, \theta_w)$ of parameters and provides a commonly used model for a prior probability density over a pmf $\mathbf{p} = (p_1, \ldots, p_{w-1})$ with a known number w of terms, which is given by:

$$f_{\mathbf{p}}(p_1, \ldots, p_{w-1}) = \frac{\Gamma(\sum_1^w \theta_i)}{\prod_{i=1}^w \Gamma(\theta_i)} \prod_i^w p_i^{\theta_i - 1},$$

$$\text{where } p_w = 1 - \sum_1^{w-1} p_i, \text{ and for } (\forall i \leq w) p_i \geq 0, \ \sum_{i=1}^{w-1} p_i \leq 1.$$

For the Dirichlet, see Fine [5] Section 7.4.3 or Wikipedia. Letting $\theta_0 = \sum_{i=1}^w \theta_i$, the means and variances $\{Ep_i, VAR(p_i)\}$ are given by

$$Ep_i = \frac{\theta_i}{\theta_0}, \ VAR(p_i) = \frac{Ep_i(1 - Ep_i)}{\theta_0 + 1}.$$

The marginal distribution of p_i is a Beta distribution given by

$$f_{p_i}(x) = \frac{\Gamma(\theta_0)}{\Gamma(\theta_i)\Gamma(\theta_0 - \theta_i)} x^{\theta_i - 1}(1 - x)^{\theta_0 - \theta_i - 1} \text{ x is in the unit interval.}$$

To preserve exchangeability, we must restrict our parameter vector θ to have all components equal to the same value, say, θ. In this case:

$$f_{\mathbf{p}}(p_1, \ldots, p_{w-1}) = \frac{\Gamma(w\theta)}{\Gamma^w(\theta)} \left(\prod_i^w p_i U(p_i) \right)^{\theta - 1}.$$

It follows from the general result above that in the exchangeable case:

$$Ep_i = \frac{1}{w}, \ VAR(p_i) = \frac{Ep_1(1 - Ep_1)}{w\theta + 1} = \frac{w - 1}{w^2(w\theta + 1)}.$$

Furthermore, it is known that in the general Dirichlet case:

$$COV(p_i, p_j) = -\frac{\theta_i \theta_j}{\theta_0^2(\theta_0 + 1)},$$

which specializes to our exchangeable case as:

$$COV(p_i, p_j) = -\frac{1}{w^2(w\theta + 1)} \text{ for } i \neq j.$$

For large w, the pairs of distinct variables are nearly uncorrelated.

As a measure of fluctuation, consider:

$$\frac{VAR(p_i)}{(EP_i)^2} = \frac{w-1}{w\theta+1} \approx \frac{1}{\theta}.$$

This is as far as our modeling has gone. We could determine θ if we could evaluate $VAR(A_i) = A^2 VAR(P_i)$ on some grounds. However, we make no determination of the remaining scalar parameter θ. The dependence between the random areas $\{A_i\}$ can make our analysis difficult. Returning to our interest in modeling the areas $\{A_i\}$, we need to rescale them to range to $[0, 1]$ through $A_i = Ap_i$. This yields:

$$EA_i = EAp_i = AEp_i = \frac{A}{w},\ VAR(A_i) = A^2 \frac{w-1}{w^2(w\theta+1)}.$$

Thus, the ratio of the standard deviation of any A_i to its mean is a fixed fraction, implying a fair degree of fluctuation in the A_i.

We will subsequently introduce approximations that require us to only know the distribution of an individual random area.

3.4 Probability of Room Identification

We are unable to determine the probability of identifying the room $\mathcal{R}$, as it requires that we be able to model the probability that all of the W walls of the room will generate Voronoi regions having a nonempty intersection with the spherical surface $S(\mathbf{0}, \underline{r})$ containing the microphones. As the Voronoi regions partition the whole space, there will be at least one Voronoi region containing the origin $\mathbf{0}$. Hence, any $S(\mathbf{0}, \underline{r})$ is intersected by at least one Voronoi region. We retrench and only ask about inference to the set $\mathcal{R}^*_{\underline{r}}$ that is the convex polyhedron generated

by just those walls that are observable by the microphone array on $S(\mathbf{0}, \underline{r})$. Of course, the observable $\mathcal{R}^*_{\underline{r}}$ contains $\mathcal{R}$ and provides an upper bound to the size of the room.

By the Poisson model, probabilities of detecting a wall are determined only by the areas $\{A_i\}$ generated and not by the actual shapes $\{S_i\}$. In the hope of gaining some insight into this problem, we arbitrarily postulate that $W^*_{\underline{r}}$ is a random variable described by a probability mass function $p_{W^*_{\underline{r}}}$, given by the Geometric distribution with parameter $0 < \gamma_{\underline{r}} < 1$:

$$P(W^*_{\underline{r}} = w) = p_{W^*_{\underline{r}}}(w) = (1 - \gamma_{\underline{r}})\gamma^{w-1}_{\underline{r}} \text{ for } w = 1, 2, \ldots. \tag{3.7}$$

As $\underline{r}$ increases, $\gamma_{\underline{r}}$ should also increase. However, we have no specific proposal to make in this regard.

For this upper bound $\mathcal{R}^*_{\underline{r}}$ to have finite volume, we need at least four walls ($W^*_{\underline{r}} \geq 4$) to form a tetrahedron. Our prior distribution over the number of "detected/observed walls" is a Geometric that starts at $w = 1$. Hence, with $P(W^*_{\underline{r}} < 4)$, we only obtain an infinite volume upper bound to the room. Consequently, the conditional probability of having a finite volume upper bound using w walls is

$$P(W^*_{\underline{r}} = w | W^*_{\underline{r}} \geq 4) = P(W^*_{\underline{r}} = w)/P(W^*_{\underline{r}} \geq 4) = (1 - \gamma_{\underline{r}})\gamma^{w-4}_{\underline{r}} \text{ for } w \geq 4, \tag{3.8}$$

and zero otherwise.

3.4.1 Conditional Probability of Identifying a Room

Let ω_i denote the event that i-th wall is identified. ω_i occurs in the noiseless case under discussion if and only if the number of sensors in A_i is at least four

(see Section 3.2.5). Let $R^*_{\underline{r}}$ denote the event that all of the observable walls are correctly identified and yield the upper bound $\mathcal{R}^*_{\underline{r}}$ to the true room. Then:

$$P(R^*_{\underline{r}}|W^*_{\underline{r}} \geq 4) = P\left(\mathcal{R}^*_{\underline{r}} = \bigcap_{i=1}^{W^*_{\underline{r}}} \omega_i \middle| W^*_{\underline{r}} \geq 4\right) = E\left\{P\left(\bigcap_{i=1}^{W^*_{\underline{r}}} \omega_i \middle| W^*_{\underline{r}}, W^*_{\underline{r}} \geq 4, \{A_1, \ldots, A_{W^*_{\underline{r}}}\}\right)\right\}. \tag{3.9}$$

The events $\{\omega_1, \ldots, \omega_W\}$ are dependent. However, we claim that they are mutually independent conditionally on knowing $W, \{A_1, \ldots, A_W\}$. The truth of this assertion follows from the Poisson point process model assumed in Section 3.3.2 and its implication that the number of sensors located in S_i, given its area A_i, is independent of the numbers of sensors located in the other Voronoi subsets of the spherical surface S and depends only upon A_i.

Therefore, by conditional independence,

$$P\left(\bigcap_{i=1}^{W^*_{\underline{r}}} \omega_i \middle| W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4, \{A_1 = a_1, \ldots, A_w = a_w\}\right) = \prod_{i=1}^{w} P(\omega_i|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4, A_i = a_i).$$

Furthermore, ω_i is independent of $W^*_{\underline{r}} = w$ if we also condition on $A_i = a_i$. Note that the enumeration of walls is not consistent from one expression to another, as the walls may or may not be observed. Instead, we are enumerating the number of partitions of $S(\mathbf{0}, \underline{r})$.

Under the Poisson point process model for sensors introduced previously on Poisson point process model for M, and using the per-unit area intensity

parameter λ for the generation of sensors, we have:

$$P(\omega_i | W_{\underline{r}}^* = w, W_{\underline{r}}^* \geq 4, A_i = a_i) \tag{3.10}$$

$$= P(\text{ at least 4 sensors in } A_i | W_{\underline{r}}^* = w, W_{\underline{r}}^* \geq 4, A_i = a_i) \tag{3.11}$$

$$= 1 - P(\text{ at most 3 sensors in } A_i | W_{\underline{r}}^* = w, W_{\underline{r}}^* \geq 4, A_i = a_i) = 1 - \phi(a_i) \text{ where} \tag{3.12}$$

$$\phi(a_i) = e^{-\lambda a_i} \left[1 + \lambda a_i + \frac{1}{2}(\lambda a_i)^2 + \frac{1}{6}(\lambda a_i)^3 \right]. \tag{3.13}$$

It follows that:

$$P\left(\bigcap_{i=1}^{W_{\underline{r}}^*} \omega_i \,\middle|\, W_{\underline{r}}^* = w, W_{\underline{r}}^* \geq 4, \{A_1 = a_1, \ldots, A_w = a_w\} \right) = \prod_{i=1}^{w} (1 - \phi(a_i)). \tag{3.14}$$

In principle, we can evaluate Equation 3.9 from Equation 3.14 by taking expectations of the latter with respect to $W_{\underline{r}}^*, A_1, \ldots, A_W$.

As discussed previously, we make the independent assumption that the random number $W_{\underline{r}}^*$ of image points having Voronoi regions that intersect with $S(\mathbf{0}, \underline{r})$ has a Geometric probability mass function and is conditioned on observable walls no less than four given by Eqn 3.8. Given $W_{\underline{r}}^* = w$, we have assumed that $A_1, \ldots, A_w$ has a Dirichlet density function as described in Section 3.3.3. We now know the joint probability law for all of these random variables.

For simplicity, we calculate the expected value appearing in Equation 3.14 by first integrating over $A_1, \ldots, A_w$, given $W_{\underline{r}}^* = w$ and $W_{\underline{r}}^* \geq 4$, and finally summing over the possible values of $W_{\underline{r}}^*$ weighted by the Poisson pmf:

$$P(R_{\underline{r}}^* | W_{\underline{r}}^* = w, W_{\underline{r}}^* \geq 4) = P\left(\bigcap_{1}^{W_{\underline{r}}^*} \omega_i \,\middle|\, W_{\underline{r}}^* = w, W_{\underline{r}}^* \geq 4 \right) = E \prod_{i=1}^{w} (1 - \phi(A_i)). \tag{3.15}$$

The difficulty at this point is that $\{A_i\}$ are mutually dependent. We will

explore some approximations to yield a tractable computation of $P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4)$.

3.4.2 Approximations to the Conditional Probability of Room Detection

We obtain an *upper bound* to $P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4)$ by recognizing that because $\phi(A)$ is a probability in $[0, 1]$

$$P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4) = E\prod_1^w(1 - \phi(A_i)) \leq 1 - \max_i E\phi(A_j) = 1 - E\phi(A_1),$$

where we use the fact that under the symmetric Dirichlet model, all of the $\{A_j\}$ are identically distributed. We can find a *lower bound* by recalling Jensen's Inequality, for a convex function f such as the exponential, that:

$$Ef(X) \geq f(EX).$$

Write:

$$E\prod_1^w(1 - \phi(A_i)) = E\exp\left\{\log\left(\prod_1^w(1 - \phi(A_i))\right)\right\} \tag{3.16}$$

$$\geq \exp\left\{E\log\left(\prod_1^w(1 - \phi(A_i))\right)\right\} = \exp\left\{E\left(\sum_1^w\log(1 - \phi(A_i))\right)\right\}. \tag{3.17}$$

We now have an expectation of a sum that is always a sum of expectations no matter how dependent the random variables are. Furthermore, by the postulated rotational symmetry that led to our choice of a symmetric Dirichlet model for the areas, each summand has the same expected value. Thus:

$$P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4) \geq \exp\left\{wE\log(1 - \phi(A_1))\right\}.$$

Combining the above results yields

$$1 - E\phi(A_1) \geq P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4) \geq \exp\{wE\log(1-\phi(A_1))\}.$$

Our estimates of $P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4)$ depend on being able to evaluate $E\phi(A_1)$ and $E\log(1-\phi(A_1))$.

As noted in Section 3.3.3, the univariate marginal of the Dirichlet is the Beta:

$$\begin{aligned} E\phi(A_1) &= \frac{\Gamma(w\theta)}{\Gamma(\theta)\Gamma((w-1)\theta)} \int_0^1 \phi(Ax) x^{\theta-1}(1-x)^{(w-1)\theta-1}\, dx \\ &= \frac{\Gamma(w\theta)}{\Gamma(\theta)\Gamma(w-1)\theta)} \int_0^1 x^{\theta-1}(1-x)^{(w-1)\theta-1} e^{-\lambda Ax}\left(1 + \lambda Ax + \frac{1}{2}\lambda^2 A^2 x^2 + \frac{1}{6}\lambda^3 A^3 x^3\right) dx. \end{aligned}$$

As we have a definite integral over a finite interval of a smooth function of a single variable, we can always evaluate it numerically. The result can then be used to find the upper bound to $P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4)$. Similarly, we can obtain $E\log(1-\phi(A_1))$ as the following:

$$E\log(1-\phi(A_1)) = \frac{\Gamma(w\theta)}{\Gamma(\theta)\Gamma((w-1)\theta)} \int_0^1 \log(1-\phi(Ax)) x^{\theta-1}(1-x)^{(w-1)\theta-1}\, dx.$$

Evaluating the integration numerically will yield the lower bound to $P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4)$.

3.4.3 Absolute Probability of Room Estimation

Now we can write:

$$P(R^*_{\underline{r}}|W^*_{\underline{r}} \geq 4) = \sum_{w=4}^{\infty} (1-\gamma_{\underline{r}})\gamma_{\underline{r}}^{(w-4)} P(R^*_{\underline{r}}|W^*_{\underline{r}} = w, W^*_{\underline{r}} \geq 4).$$

As we cannot identify a room with w walls if we have fewer than $4w$ sensor points, at the least, we require:

$$E|M| = \lambda A = 4\pi(\underline{r})^2\lambda > 4EW = 4\gamma_{\underline{r}}.$$

3.5 MATLAB Simulation

3.5.1 Simulation Setup

The MATLAB simulation in this sub-section is to show the distribution of 5 microphones/30 microphones onto a unit circle contained in a 2D square room:

1. Given a 2D unknown square room of size 4×4;
2. The vertices of the room are at $(2, 2), (2, -2), (-2, 2)$, and $(-2, -2)$;
3. The sound source S is located at the origin;
4. The circle that we restrict our microphones on has a unit radius;
5. Let the microphones be uniformly distributed on the circle.

Then the result of distributing five microphones is in Fig 3.1, and that of distributing 30 microphones is in Fig 3.2. Clearly, when the number of microphones increases, the likelihood of having four or more microphones in each Voronoi region and recovering the images also increases.

3.5.2 Detecting the Room

This simulation is to show the relationship between the number of microphones and the number of detectable walls when the microphones are randomly distributed. The setup is as follows:

1. Given a 3D cubical room of size $4 \times 4 \times 4$;

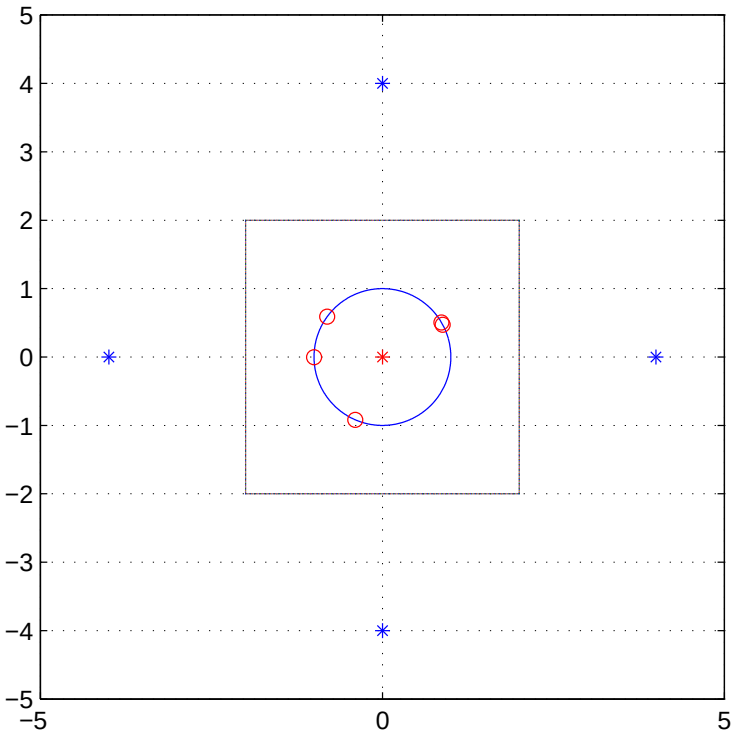


Figure 3.1: Distributing five microphones on a unit circle

2. The sound source S is located at the origin;
3. The sphere that we restrict our microphones on has a unit radius;
4. Uniformly distribute microphones on the sphere;
5. For each trial of distributing microphones, obtain the number of detectable walls;
6. For each number of K microphones, run the trial for 50 times and obtain the number of times that w number of walls were detected where w=1..6.

The following table captures the result of the above experiment:

K value w=1 w=2 w=3 w=4 w=5 w=6

K=4	0	0	0	0	0	0
K=5	1	0	0	0	0	0
K=6	2	0	0	0	0	0
K=7	7	0	0	0	0	0
K=8	5	0	0	0	0	0
K=9	17	0	0	0	0	0
K=10	23	2	0	0	0	0
K=11	26	3	0	0	0	0
K=12	27	4	0	0	0	0
K=13	29	6	0	0	0	0
K=14	27	14	0	0	0	0
K=15	20	26	0	0	0	0
K=16	14	31	2	0	0	0
K=17	10	27	13	0	0	0
K=18	10	24	16	0	0	0
K=19	3	25	21	1	0	0
K=20	4	19	22	5	0	0
K=21	3	11	25	10	1	0
K=22	0	14	30	6	0	0
K=23	0	6	19	22	3	0
K=24	0	3	25	18	4	0
K=25	0	6	18	19	7	0
K=26	0	2	15	24	9	0
K=27	0	0	10	27	12	1
K=28	0	1	9	21	17	2
K=29	0	0	5	21	23	1

K=30	0	0	7	13	19	11
K=31	0	0	3	17	26	4
K=32	0	0	2	9	27	12
K=33	0	0	1	12	26	11
K=34	0	0	3	13	19	15
K=35	0	0	0	6	32	12
K=36	0	0	1	6	22	21
K=37	0	0	0	9	17	24
K=38	0	0	0	5	24	21
K=39	0	0	0	1	12	37
K=40	0	0	0	2	25	23
K=41	0	0	0	0	16	34
K=42	0	0	0	3	16	31
K=43	0	0	0	0	12	38
K=44	0	0	0	0	10	40
K=45	0	0	0	1	8	41
K=46	0	0	0	1	14	35
K=47	0	0	0	0	9	41
K=48	0	0	0	1	8	41
K=49	0	0	0	1	6	43
K=50	0	0	0	0	10	40

Table 3.1: Table of number of identifiable walls result

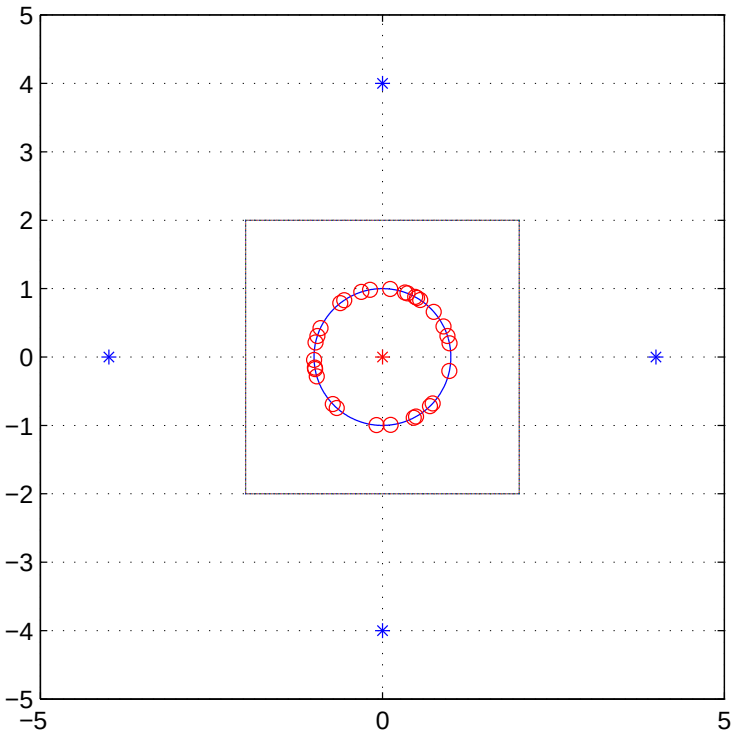


Figure 3.2: Distributing 30 microphones on a unit circle

The above table illustrated the odd of recovering W walls given K microphones deployed. As the number of microphone increases, the odds of obtaining more walls is increased. Heuristically from the above simulation, to recover six walls with more than fifty-percent chances, we would need more than 40 microphones deployed. Although this seems a large number of microphones as compared with four microphones per wall, thus 24 microphones are required, we make a complete random case here, that is, the room is random, and microphones are randomly distributed. These assumptions need more microphones to have at least four microphones closest to each of the walls.

3.6 Conclusion

In conclusion, this Chapter provides system models for partially identifying an unknown room from noise-free wave measurements. Based on that model, we analyze the probability of estimating an upper bound $\mathcal{R}^*_{\underline{r}}$ to the true room $\mathcal{R}$.

CHAPTER 4

LOCALIZATION OF AN IMAGE SOURCE UNDER NOISE

Recall that in the noiseless case in Chapter 3, the information gained from the first echo received at a single microphone located at $\mathbf{m}$ establishes the locus of the closest image point $\mathbf{I}$ as the spherical *surface* $S(\mathbf{m}, \rho)$ of center $\mathbf{m}$ and radius ρ, the distance between $\mathbf{m}$ and $\mathbf{I}$. Thus, localizing $\mathbf{I}$ becomes easy when there are four or more distinct microphone locations closest to the same image point, as in Section 3.2; this image point could be uniquely identified as:

$$\mathbf{I} = \bigcap_{i=1}^{M} S(\mathbf{m}_i, \rho_i), \ M \geq 4.$$

In the noisy case, however, the distance measurement does not represent the true distance between a microphone and its closest image. Therefore, the intersection of $S(\mathbf{m}_i, \rho_i)$'s does not always have a unique point, and we are not able to determine the image location. In this chapter, we are going to first formulate the localization problem to estimate the image location based on noisy distance measurements and then give a solution that corresponds to our assumptions.

4.1 Introduction

Finding the image source based on distance measurements is usually referred to as a localization problem, a classical range-based localization problem in sensor networks. The problem based on distance measurements usually leads to a non-linear optimization. A number of solutions have already been developed in mainly two categories: iterative processes (see Foy [6]) and closed form solutions (see Zhao and Guibas [16], Manolaskis [12]).

The iterative process to solve a non-linear localization problem is usually initiated by a Taylor Series expansion of the original problem, keeping only the lower order terms, and then iteratively by Newton's method search for the optimal point based on an initial guess (see Foy [6]). One of the major disadvantages of this technique is that the solution requires extensive computation to obtain the Hessian matrix for Taylor expansion, as most of the time that matrix is not in closed form. Another disadvantage is that a "smart" guess is required to start the iterative process. Depending on the initial guess, the method may not converge to the correct value. The closed form solution for range-based localization, on the other hand, usually requires linearizing the problem and estimating the optimal value for a linearized problem (see Zhao and Guibas [16], Manolaskis [12]). However, it is questionable how "far" is the linearized problem from the original problem.

In this dissertation, based on the applications we have in mind, the localization problem is formulated as solving for the least squared error. The cost function was designed for an iterative process, which provides one distinct feature: a closed form Hessian matrix that significantly reduces the computational complexity.

4.1.1 Assumptions

Before getting into the details, we are going to assume the following:

- We have K omnidirectional microphones located at $M = \{\mathbf{m}_1, ..., \mathbf{m}_K\}$ lying on the same side of a wall as the sound source as $\mathbf{0}$.
- The locations of microphones are non-coplanar, that is, no four or more

microphones are on a plane.

- The image location **I** is unknown but assumed to be within a known finite domain $\mathcal{D}$.
- Image **I** is closest to microphones in set M, and the distances ρ_i between $\mathbf{m}_i$ and **I** are estimated using the process from Chapter 2.

4.1.2 Noise Model

The noise $n(t)$ considered in this chapter is the same as described in Section 1.2.5, and with synchronization noise detailed as follows:

Noise due to synchronization: To model the effect of timing errors, we adopt a model similar to that considered in Hu and Servetto [9][10], which has been previously verified empirically in Elson, Girod and Estrin [4]: Every time a clock is queried, it produces a reading corrupted by Gaussian noise of a fixed variance, and the noise is independent across multiple queries. In the context of our wave readings, this means that if $f(\mathbf{m}_i)$ is a reading of the wave at location $\mathbf{m}_i$ and at true time t, then when the sensor queries his clock, the clock will return a value:

$$\hat{t} \sim \mathcal{N}(t, \sigma^2). \tag{4.1}$$

As a result, when the central computing unit obtains the time of arrival $\hat{t}$ of the echo according to the master clock and estimates the distance from the sample to the nearest image point as $c\hat{t}$, this reading takes the form:

$$\hat{\rho} \;=\; c\hat{t} \;=\; c(t+e) \;=\; \rho + ce, \tag{4.2}$$

where

- c is the speed of wave propagation;
- $e \sim \mathcal{N}(0, \sigma^2)$ is the measurement noise;
- t is the actual time at which the second peak is seen;
- ρ is the true distance between $\mathbf{m}_i$ and the closest image.

So the noisy distance measurement by microphone $\mathbf{m}_i$ can be written as:

$$\rho_i = \rho_i^0 + e_i,$$

where ρ_i^0 is the unknown true distance, and e_i is the error term that includes the noise as described above.

4.2 Reconstruction of an Image from Noisy Distance Estimates

4.2.1 Problem Formulation

Based on the assumptions in this chapter, the true distance between microphone $\mathbf{m}_i$ and image $\mathbf{I}$ is $|\mathbf{m}_i - \mathbf{I}|$; so the sum of the squared error can be written as:

$$C'(\mathbf{I}) = \sum_{i=1}^{K} \left(\rho_i^2 - |\mathbf{m}_i - \mathbf{I}|^2\right)^2. \tag{4.3}$$

We want to choose $\mathbf{I}$ to minimize $C'(\mathbf{I})$. Note that C' is used to differentiate that this cost function is based on one signal measurement of ρ_i, and using multiple measurements is addressed immediately following this sub-section. Here, we use the difference of the squared values for cost instead of using the conventional sum of the squared error $\sum_{i=1}^{K} (\rho_i - |\mathbf{m}_i - \mathbf{I}|)^2$. The conventional sum of the squared error has been studied in various places (see Zhao and Guibas [16],

Zhao and Servetto [17]). However, due to its high non-convexity, the solution to it usually becomes either over-complicated or over-simplified. In this dissertation, we use the alternative sum of the squared error so that we can provide a reasonable solution without trivializing the problem.

4.2.2 Using the Average Distance Measurement

Now let's have n repeated independent and identically distributed measurements $\{\rho_i^{(j)}\}$ of ρ_i; the total cost function that we want to minimize becomes:

$$C(\mathbf{I}) = \sum_{j=1}^{n} \sum_{i=1}^{K} \left((\rho_i^{(j)})^2 - |\mathbf{m}_i - \mathbf{I}|^2 \right)^2 . \tag{4.4}$$

Let:

$$\bar{\rho_i^2} = \frac{1}{n} \sum_{j=1}^{n} (\rho_i^{(j)})^2 .$$

Then expanding and re-configuring Eqn. 4.4 yields:

$$C(\mathbf{I}) = \sum_{j=1}^{n} \sum_{i=1}^{K} \left(\left((\rho_i^{(j)})^2 \right)^2 - 2(\rho_i^{(j)})^2 |\mathbf{m}_i - \mathbf{I}|^2 + \left(|\mathbf{m}_i - \mathbf{I}|^2 \right)^2 \right) \tag{4.5}$$

$$= \sum_{i=1}^{K} \left(\sum_{j=1}^{n} \left((\rho_i^{(j)})^2 \right)^2 - 2 \sum_{j=1}^{n} (\rho_i^{(j)})^2 |\mathbf{m}_i - \mathbf{I}|^2 + n \left(|\mathbf{m}_i - \mathbf{I}|^2 \right)^2 \right) \tag{4.6}$$

$$= n \sum_{i=1}^{K} \left(\frac{1}{n} \sum_{j=1}^{n} \left((\rho_i^{(j)})^2 \right)^2 - 2\bar{\rho_i^2} |\mathbf{m}_i - \mathbf{I}|^2 + \left(|\mathbf{m}_i - \mathbf{I}|^2 \right)^2 \right) \tag{4.7}$$

$$= n \sum_{i=1}^{K} \left(\frac{1}{n} \sum_{j=1}^{n} \left((\rho_i^{(j)})^2 \right)^2 + \left(\bar{\rho_i^2} \right)^2 - \left(\bar{\rho_i^2} \right)^2 - 2\bar{\rho_i^2} |\mathbf{m}_i - \mathbf{I}|^2 + \left(|\mathbf{m}_i - \mathbf{I}|^2 \right)^2 \right) \tag{4.8}$$

$$= n \sum_{i=1}^{K} \left(\frac{1}{n} \sum_{j=1}^{n} \left((\rho_i^{(j)})^2 \right)^2 - \left(\bar{\rho_i^2} \right)^2 + \left(\bar{\rho_i^2} - |\mathbf{m}_i - \mathbf{I}|^2 \right)^2 \right) \tag{4.9}$$

$$= n \sum_{i=1}^{K} \left(\frac{1}{n} \sum_{j=1}^{n} \left((\rho_i^{(j)})^2 \right)^2 - \left(\bar{\rho_i^2} \right)^2 \right) + n \sum_{i=1}^{K} \left(\left(\bar{\rho_i^2} - |\mathbf{m}_i - \mathbf{I}|^2 \right)^2 \right). \tag{4.10}$$

Ignoring the terms that are not dependent on $\mathbf{I}$ in Eqn.4.10, we have that minimizing Eqn. 4.4 is equivalent to minimizing the equation:

$$C(\mathbf{I}) = \sum_{\mathbf{i=1}}^{\mathbf{K}} \left(\left(\bar{\rho_{\mathbf{i}}^2} - |\mathbf{m_i} - \mathbf{I}|^{\mathbf{2}} \right)^{\mathbf{2}} \right). \tag{4.11}$$

Now, we are going to estimate the image location by minimizing Eqn. 4.11.

4.2.3 Taylor Series Expansion of the Cost Function

Assuming row vectors, the cost function in Eqn. 4.11 can be written as:

$$\begin{aligned} C(\mathbf{I}) &= \sum_{i=1}^{K} \left(\bar{\rho_i^2} - |\mathbf{m_i} - \mathbf{I}|^{\mathbf{2}} \right)^2 \\ &= \sum_{i=1}^{K} \left(\bar{\rho_i^2} - |\mathbf{m_i}|^{\mathbf{2}} - |\mathbf{I}|^{\mathbf{2}} + \mathbf{2 m_i I^T} \right)^2 . \end{aligned}$$

Introducing a small perturbation $\delta\mathbf{I}$ around $\mathbf{I}$ yields:

$$\begin{aligned} C(\mathbf{I} + \delta\mathbf{I}) &= \sum_{i=1}^{K} \left(\bar{\rho_i^2} - |\mathbf{m_i}|^{\mathbf{2}} - |\mathbf{I} + \delta\mathbf{I}|^{\mathbf{2}} + \mathbf{2 m_i (I + \delta I)^T} \right)^2 \\ &= \sum_{i=1}^{K} \left(\bar{\rho_i^2} - |\mathbf{m_i}|^{\mathbf{2}} - |\mathbf{I}|^{\mathbf{2}} - |\delta\mathbf{I}|^{\mathbf{2}} - \mathbf{2 I \delta I^T} + \mathbf{2 m_i I^T} + \mathbf{2 m_i \delta I^T} \right)^2 \\ &= \sum_{i=1}^{K} \left(\bar{\rho_i^2} - |\mathbf{m_i}|^{\mathbf{2}} - |\mathbf{I}|^{\mathbf{2}} + \mathbf{2 m_i I^T} - |\delta\mathbf{I}|^{\mathbf{2}} - \mathbf{2 I \delta I^T} + \mathbf{2 m_i \delta I^T} \right)^2 \\ &= \sum_{i=1}^{K} (A_i - B_i)^2 , \end{aligned}$$

where:

$$A_i = \bar{\rho_i^2} - |\mathbf{m_i}|^{\mathbf{2}} - |\mathbf{I}|^{\mathbf{2}} + \mathbf{2 m_i I^T}$$

$$B_i = |\delta\mathbf{I}|^{\mathbf{2}} + \mathbf{2 I \delta I^T} - \mathbf{2 m_i \delta I^T}.$$

Therefore:

$$\begin{aligned}
C(\mathbf{I}+\delta\mathbf{I}) &= \sum_{i=1}^{K}\left(A_i^2 - 2A_iB_i + B_i^2\right) \\
&= \sum_{i=1}^{K} A_i^2 - 2\sum_{i=1}^{K} A_i\left(|\delta\mathbf{I}|^{\mathbf{2}} + \mathbf{2I}\delta\mathbf{I^T} - \mathbf{2m_i}\delta\mathbf{I^T}\right) \\
&+ \sum_{i=1}^{K}\left(|\delta\mathbf{I}|^{\mathbf{2}} + \mathbf{2I}\delta\mathbf{I^T} - \mathbf{2m_i}\delta\mathbf{I^T}\right)^2 \\
&= \sum_{i=1}^{K} A_i^2 + 4\sum_{i=1}^{K} A_i(\mathbf{m_i}-\mathbf{I})\delta\mathbf{I}^T + 2\sum_{i=1}^{K}\delta\mathbf{I}\left(2(\mathbf{m_i}-\mathbf{I})^T(\mathbf{m_i}-\mathbf{I}) - \mathbb{I}A_i\right)\delta\mathbf{I^T} \\
&+ O(|\delta\mathbf{I}|^{\mathbf{3}}),
\end{aligned}$$

where $\mathbb{I}$ is the identity matrix to match the dimension. Ignoring the $O(|\delta\mathbf{I}|^{\mathbf{3}})$ terms yields:

$$C(\mathbf{I}+\delta\mathbf{I}) = \sum_{i=1}^{K} A_i^2 + 4\sum_{i=1}^{K} A_i(\mathbf{m_i}-\mathbf{I})\delta\mathbf{I}^T + 2\sum_{i=1}^{K}\delta\mathbf{I}\left(2(\mathbf{m_i}-\mathbf{I})^T(\mathbf{m_i}-\mathbf{I}) - \mathbb{I}A_i\right)\delta\mathbf{I}^T. \quad (4.12)$$

Then matching the coefficient of the quadratic form of cost function to the Taylor series expansion of $C(\mathbf{I}+\delta\mathbf{I})$ around $\mathbf{I}$, we can obtain that the gradient is:

$$\nabla_{\mathbf{I}} C = 4\sum_{i=1}^{K} A_i(\mathbf{m_i}-\mathbf{I}). \quad (4.13)$$

And the Hessian matrix is:

$$\mathbb{H}_{\mathbf{I}} = 4\sum_{i=1}^{K}\left(2(\mathbf{m_i}-\mathbf{I})^T(\mathbf{m_i}-\mathbf{I}) - \mathbb{I}A_i\right). \quad (4.14)$$

4.2.4 Algorithm for the Least Squares Problem

With the gradient and Hessian matrix, we are able to solve the least squares problem by using Newton's method. The algorithm is as follows:

1. Start with a initial guess of $\mathbf{I^{(1)}}$ that yields a positive definite Hessian.

2. Apply Newton's method with starting point $\mathbf{I^{(1)}}$ and repeat with:

$$\mathbf{I^{(n+1)}} = \mathbf{I^{(n)}} - \nabla_{\mathbf{I^{(n)}}}\mathbf{C}[\mathbb{H}_{\mathbf{I^{(n)}}}]^{-1}.$$

3. Stop the process when Newton's method terminates with a previously chosen step size ϵ, e.g., $\epsilon = 0.0001$, difference between successive iterations.

4. Generate the estimate for $\mathbf{I}$.

4.2.5 Analysis

The advantages of this Newton's algorithm approach include:

- Easy to calculate, as the gradient and Hessian matrix have a closed form based on the distance measurements and the optimal value from previous iteration. Thus, there is no costly computation for obtaining the Hessian matrix.
- The convergence is fast as a result of Newton's method.
- The algorithm doesn't depend on dimension. It can be applied directly to the 2D or 3D cases.
- The algorithm stops at the choice of step size ϵ in Newton's method.

On the other hand, there is one disadvantage:

- It requires an initial guess, and if the guess is wrong, the algorithm may not converge or may converge to some suboptimal point that is far from the true location of the image.

This disadvantage, however, could be mitigated by choosing the initial point that is close to the the true value. For example, one can choose a starting point that is one of the spherical surfaces' intersections; then if the noise is small, it is very likely that the true image will be surrounded by the intersections. This way it is more likely to obtain a good estimate.

4.3 MATLAB Simulation

4.3.1 Simulation Setup

We ran a simulation of localizing an image point based on noisy distance measurements. The four microphones were located at:

$$\mathbf{m_1} = (0, 0, 10),$$

$$\mathbf{m_2} = (0, 10, 0),$$

$$\mathbf{m_3} = (10, 0, 0),$$

$$\mathbf{m_4} = (5, 5, -5).$$

The image is located at $\mathbf{I} = (10, 8, -2)$, which is unknown. The true distances are:

$$\rho_1^0 = 17.5499,$$

$$\rho_2^0 = 10.3923,$$

$$\rho_3^0 = 8.2462,$$

$$\rho_4^0 = 6.5574.$$

We add zero mean Gaussian noise with variance one to each of the four distance measurements. So the noisy distance measurements are:

$$\rho_1 = 18.5000,$$

$$\rho_2 = 10.6234,$$

$$\rho_3 = 8.8530,$$

$$\rho_4 = 7.0434.$$

4.3.2 Choosing the Initial Point

Assuming the noise involved in obtaining distance measurement is small, as a result of increasing probe signal, we expect that although the four spheres do not intersect at one single point, the intersections of every three spheres will cluster around the true image. Therefore, choosing an intersection of any of the three spheres as an initial point for Newton's method would likely yield an initial estimate closer to the true value. Starting with $\mathbf{m_1}, \mathbf{m_2}, \mathbf{m_3}$ and ρ_1, ρ_2, ρ_3 to start with, we obtain two points: $i_1 = (6.6432, 4.9191, -6.5505)$ and $i_2 = (9.9687, 8.2446, -3.2251)$. Those two points are the initial points we use for Newton's methods.

4.3.3 Newton's Method

Applying Newton's method starting from i_1 yields an estimate of the image location: $\hat{\mathbf{I}}_1 = (6.1288, 4.0757, -6.9921)$; and applying Newton's method starting from i_2 yields an estimate of the image location: $\hat{\mathbf{I}}_2 = (10.3362, 8.6278, -2.6227)$.

Plugging $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$ in Eqn.5.3 yields the squared error:

$$C(\hat{\mathbf{I}}_1) = 1973.8 \;\; C(\hat{\mathbf{I}}_2) = 24.9364.$$

Thus, we choose the one with the least squared error. The estimate is $(10.3362, 8.6278, -2.6227)$, which is very close to $(10, 8, -2)$ the true value with an error of 0.946 measured by L^2-norm.

4.4 Conclusion

In this chapter, we have formulated the problem of localizing an image point based on noisy distance measurements as a least squares estimation problem. An iterative solution to the least squares estimation is provided based on Newton's method and a MATLAB simulation is performed. The solution introduced in this chapter has some key features that in our application make significant contributions:

- The least squares estimation does not require iteratively taking measurements of distances.
- The solution introduced here is based on a closed form Hessian matrix, and thus, simply applying Newton's method obtains the result; there is no complexity spent in approximating the Hessian matrix or its inverse.
- When the noise in a distance measurement is small, the initial point for Newton's method can be chosen very close to the true image location. Thus, even though the solution may not be a global minimum to the least squares problem, it may be close to the true value.

CHAPTER 5

IDENTIFICATION OF AN UNKNOWN ROOM WITH NOISE

Unlike in Chapter 3, the algorithm to identify an unknown room from noisy measurements is heavily dependent on the particular application and the flexibility of how much we know about the room. For example, the algorithm for a robot to discover the shape of a coal mine is more complicated than the algorithm for a home entertainment system to recover the shape of a room where one can access the number of walls and place the microphones close to known boundaries. However, what we have in mind does not have any prior information on the unknown room. In this chapter, we utilize the results from the previous two chapters to identify an unknown room under noisy measurements.

5.1 Room Identification under Noise

The steps that we take to recover room $\mathcal{R}$ are as follows:

1. Estimate the distance between each microphone and its closest image.
2. Identify the cluster of microphones that are closest to each image.
3. Estimate the image sources and the corresponding locations.
4. Recover the room.

5.1.1 Distance Measurement from the Wave Sample

As in the noise-free case, the method of inference to the walls under noise is by estimating the delay to the first echo as measured at each $\mathbf{m}_i$ obtained from the microphone wave sample. This delay, combined with the known speed c of sound, enables us to determine the *scalar distance* ρ_i from $\mathbf{m}_i$ to the image point $\mathbf{I}_{i^*}$ that is closest to $\mathbf{m}_i$. Let the wave measurement at microphone $\mathbf{m}_i$ be $r_i(t)$, described in Section 2.2.2, the distance ρ_i can be estimated as:

$$\hat{\rho}_i = \arg\min_{\delta \in \Delta} \| r_{i,2}(t) - \frac{A}{\delta + d_1} s(t - \frac{d_1 + \delta}{c}) \| \tag{5.1}$$

where

$$r_{i,2}(t) = r_i(t) 1_{[\frac{d_1}{c} + T^*, T]}(t). \tag{5.2}$$

We can repeat this procedure on multiple wave measurements at $\mathbf{m}_i$ and obtain the sample square mean $\overline{\rho^2}_i$ of $\hat{\rho}_i$. By the analysis in Chapter 2, in a given environment, increasing the amplitude of the probe signal $s(t)$ generated from the source can reduce the error of $\overline{\rho_i}$ (and $\overline{\rho^2}_i$) from the true value.

5.1.2 Identifying Possible Clusters

To identify the cluster of microphones, we have to first know which microphones are closest to the same image point. In the noise-free case, the decision is made based on the quadruple of spherical surfaces generated by four distinct microphones: If the quadruple of spherical surfaces,

$$(S(\mathbf{m}_i, \rho_i), S(\mathbf{m}_j, \rho_j), S(\mathbf{m}_k, \rho_k), S(\mathbf{m}_l, \rho_l)),$$

has a single intersection, then the microphones $\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k$ and $\mathbf{m}_l$ are closest to a common image point and, thus, in the same cluster. In the noisy case, however,

the quadruple of spherical surfaces,

$$(S(\mathbf{m}_i, \bar{\rho}_i), S(\mathbf{m}_j, \bar{\rho}_j), S(\mathbf{m}_k, \bar{\rho}_k), S(\mathbf{m}_l, \bar{\rho}_l)),$$

will usually not intersect at one common point, as the noisy distance estimates $\bar{\rho}_i, \bar{\rho}_j, \bar{\rho}_k$, and $\bar{\rho}_l$ in general do not equal to the true distances. In the succeeding discussion, we are going to obtain an estimate of a possible image point based on the four spherical surfaces.

Notation:

1. Let $\mathbf{M} = \{\mathbf{m_i}\}$, with $M = \|\mathbf{M}\|$, denote the known set of sensor locations written as row vectors.
2. A variable under a tilde sign is thought of as a true, but sometimes unknown, quantity.
3. Let $\tilde{\mathbf{I}} = \{\mathbf{I}\}$ denote the true, but unknown, set of image points, written as row vectors.
4. Define,

$$\tilde{\rho}_i = \min_{\mathbf{I} \in \tilde{\mathbf{I}}} |\mathbf{I} - \mathbf{m_i}|, \tag{5.3}$$

   as the true, but unknown, distance between sensor $\mathbf{m}_i$ and its closest image point in the true set of image points $\tilde{\mathbf{I}}$. Define $\bar{\rho}_i$ as the estimated value of $\tilde{\rho}_i$ when we have n repeated independent and identically distributed measurements $\{\rho_i^{(j)}\}$ of $\tilde{\rho}_i$.
5. Define a true set of quadruples:

$$\bar{Q}(i_1, i_2, i_3, i_4) = \{(\tilde{\rho}_{i_k}, \mathbf{m}_{i_k}) : k = 1 : 4\}. \tag{5.4}$$

   And an estimated set of quadruples:

$$\bar{Q}(i_1, i_2, i_3, i_4) = \{(\bar{\rho}_{i_k}, \mathbf{m}_{i_k}) : k = 1 : 4\}. \tag{5.5}$$

Our goal is to determine whether or not $\bar{Q}(i_1, i_2, i_3, i_4)$ corresponds to a single true image point that we can estimate. If so, then we say that the sensor locations are clustered together. Knowing this clustering we can then estimate the image point location. In what follows we will generically take $\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2$, and $\mathbf{m}_3$ to denote a quadruple of sensor sites under consideration. The need for a certain matrix to be invertible will require that these four sensors are not coplanar.

Solving for the Purported Image Point: Noiseless Case

By "purported," we mean a common closest image point to the four sensors, if such exists. In the noiseless case, when the four sensors cluster together in that they are all closest to the same image point, then the image point must satisfy:

$$(\forall 0 \leq i \leq 3)|\mathbf{m}_i - \mathbf{I}| = \tilde{\rho}_i \text{ or, equivalently, } |\mathbf{m}_i - \mathbf{I}|^2 = \tilde{\rho}_i^2. \tag{5.6}$$

Expanding, we see that:

$$|\mathbf{m}_i - \mathbf{I}|^2 = |\mathbf{m}_i|^2 + |\mathbf{I}|^2 - 2\mathbf{m}_i \cdot \mathbf{I} = \tilde{\rho}_i^2. \tag{5.7}$$

This equation is quadratic in $|\mathbf{I}|$, but it can be transformed into an equation linear in $\mathbf{I}$ by subtraction.

$$|\mathbf{m}_i - \mathbf{I}|^2 - |\mathbf{m}_0 - \mathbf{I}|^2 = |\mathbf{m}_i|^2 - |\mathbf{m}_0|^2 - 2(\mathbf{m}_i - \mathbf{m}_0) \cdot \mathbf{I} = \tilde{\rho}_i^2 - \tilde{\rho}_0^2. \tag{5.8}$$

Let $\mathbf{m}_i^0 = \mathbf{m}_i - \mathbf{m}_0$ and $\mathbf{I}^0 = \mathbf{I} - \mathbf{m}_0$. Equivalently, we shift the origin of the coordinate system from $\mathbf{0}$ to $\mathbf{m}_0$ and obtain:

$$(\forall 1 \leq i \leq 3)\ 2\mathbf{m}_i^0 \cdot \mathbf{I}^0 = |\mathbf{m}_i|^2 - |\mathbf{m}_0|^2 + \tilde{\rho}_0^2 - \tilde{\rho}_i^2. \tag{5.9}$$

Hence, letting,

$$\mathbb{M}_0 = \begin{pmatrix} \mathbf{m}_1^0 \\ \mathbf{m}_2^0 \\ \mathbf{m}_3^0 \end{pmatrix} \text{ and } \mathbf{c} = \begin{pmatrix} |\mathbf{m}_1|^2 - |\mathbf{m}_0|^2 + \tilde{\rho}_0^2 - \tilde{\rho}_1^2 \\ |\mathbf{m}_2|^2 - |\mathbf{m}_0|^2 + \tilde{\rho}_0^2 - \tilde{\rho}_2^2 \\ |\mathbf{m}_3|^2 - |\mathbf{m}_0|^2 + \tilde{\rho}_0^2 - \tilde{\rho}_3^2 \end{pmatrix}. \tag{5.10}$$

We have:

$$\mathbb{M}_0 \mathbf{I}^0 = \frac{1}{2}\mathbf{c}. \tag{5.11}$$

Assuming that the quadruple of sensor locations are not coplanar, the inverse matrix exists, and we have the solution:

$$\mathbf{I}^0 = \frac{1}{2}\mathbb{M}_0^{-1}\mathbf{c}. \tag{5.12}$$

Hence, we can solve for the image point $\mathbf{I} = \mathbf{I}^0 + \mathbf{m}_0$ corresponding to the cluster of four sensors.

However, the four sensors need not be in the same cluster. To determine whether this is the case, we return to the condition for sensor location $\mathbf{m}_0$ and check to see if:

$$|\mathbf{m}_0 - \mathbf{I}|^2 = |\mathbf{I}^0|^2 = \tilde{\rho}_0^2 \text{ or if } |\mathbf{I}^0| = \tilde{\rho}_0. \tag{5.13}$$

The sensors at the four locations have a common closest image point if and only if Eqn. 5.13 holds.

Solving for the Purported Image Point: Noisy Case

By repeated emission of n pulses from the signal source, we can generate independent and identically distributed measurements $\{\rho_i^{(j)}\}$ of the distance from sensor $\mathbf{m}_i$ to its nearest image source. The most straightforward approach is to solve for $\{\mathbf{I}_j^0\}$ by using the j-th measurements and Eq. 5.12. The matrix $\mathbb{M}_0$ is unchanged in successive measurements. Hence, we require only one evaluation of its inverse. The vector $\mathbf{c}$ is now a set $\{\mathbf{c}_j\}$ of values due to the changing values of the squares of the sensor-image distance measurements.

By reference to Eq. 5.13, our interest in clustering requires us to also determine $|\mathbf{I}^0|$. We use standard statistical techniques of least mean-square unbiased

linear estimators of the means:

$$\hat{m}_{|\mathbb{I}|} = \frac{1}{n}\sum_{j=1}^{n} |\mathbf{I}_j^0|, \quad \hat{m}_{\mathbf{I}^0} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{I}_j^0. \tag{5.14}$$

Similarly, we can estimate the variance of this estimator from:

$$\hat{\sigma}^2_{\hat{m}_{|\mathbb{I}|}} = \frac{1}{n-1}\sum_{j=1}^{n} \{|\mathbf{I}_j^0| - \hat{m}_{|\mathbb{I}|}\}^2. \tag{5.15}$$

With these two results, we can define an approximate (it uses estimated rather than true quantities) confidence interval of, say, plus or minus τ standard deviations about the mean:

$$L_{|I^0|} = \left(\hat{m}_{|\mathbb{I}|} - \tau\hat{\sigma}_{\hat{m}_{|\mathbb{I}|}}, \hat{m}_{|\mathbb{I}|} + \tau\hat{\sigma}_{\hat{m}_{|\mathbb{I}|}}\right). \tag{5.16}$$

If the errors in $|\mathbf{I}^0|$ had a normal distribution, then the interval $L_{|I^0|}$, for $\tau = 2$, would contain the true value with probability of about 0.95 for n large enough that the estimated mean and variance are accurate.

We now repeat the above estimation for $\tilde{\rho}_0$, yielding a sample mean $\hat{m}_{\rho_0}$ and a sample variance $\hat{\sigma}^2_{\rho_0}$. We can then define a plus or minus τ sigma confidence interval for $\tilde{\rho}_0$:

$$L_{\rho_0} = \left(\hat{m}_{\rho_0} - \tau\hat{\sigma}_{\rho_0}, \hat{m}_{\rho_0} + \tau\hat{\sigma}_{\rho_0}\right). \tag{5.17}$$

We distinguish three disjoint and exhaustive alternatives:

1. For a given τ_1, if the two confidence intervals overlap, then we decide that $|\mathbf{m}_{\mathbb{I}^0} + \mathbf{m}_0| = \tilde{\rho}_0$ and, therefore, that this quadruple of sensors does form a cluster with $\hat{m}_{\mathbb{I}}$ as its closest common image point.
2. If for a given τ_2, the two confidence intervals are disjoint, then we decide that $|\mathbf{m}_{\mathbb{I}^0} + \mathbf{m}_0| \neq \tilde{\rho}_0$ and the quadruple of sensors does not form a cluster. In this case, we go on to consider another quadruple.

3. If, for the given τ_1, τ_2, neither of the preceding two cases holds, then we are unable to reach a decision and require additional measurements to resolve this impasse.

The methods based on *Wald's sequential analysis* can provide an organized approach to resolving statistical questions in which a variable amount of data is needed to securely choose between two alternatives. However, we do not pursue this line here. For more information on the sequential analysis, please refer to Sen [13] and Siegmund [14].

It is important to notice that when there is no noise, Eqn. 5.16 and Eqn. 5.17 are two points. This is equivalent to checking the intersection of quadruple spheres described in Chapter 3. When there is large noise in the measurements, Eqn. 5.16 and Eqn. 5.17 could be large intervals, and we may group sensors that do not belong to the same cluster. Therefore, we may want to take the average of more measurements to obtain a better clustering of the quadruples.

Finally, having examined all quadruples, we likely will need to further cluster the resulting image points and their associated closest sensors. Given that we are estimating the image points, the individual estimation errors will almost surely yield different image points that in reality correspond to small estimation errors made on a common image point. It would seem that this further clustering could be guided by the clique algorithm and by determining whether or not they overlap.

5.1.3 An Ad-Hoc Approach to Inferring Image Points in the Noisy Case

Setup

1. First, for given τ_1 and τ_2, let set Q_1, Q_2, and Q_3 be the sets of sensor quadruples that fall in alternative 1, alternative 2, and alternative 3, respectively, from the previous section.
2. Then Q_1 contains the quadruples of sensors that have a common image point with approximate confidence interval defined by τ_1. Q_2 contains the quadruples of sensors that have no common image point with approximate confidence interval defined by τ_2. And Q_3 contains the quadruples of sensors for which we cannot identify whether they have a common image point or not without making additional measurements. Additional measurements may eliminate the quadruple in Q_3; however, we are not making such a claim here. And we restrict our further clustering to Q_1.
3. Our approach to further clustering sensors is by the number of cliques in a connected graph.
4. Denote as $\mathbf{M}^*$ the subset of $\mathbf{M}$ consisting of sensor locations in Q_1, and let K be the number of elements in $\mathbf{M}^*$.
5. We then define an undirected graph $\mathcal{G}$ on vertices $1, ..., K$ by making i and j adjacent if and only if m_i and m_j lie together in some quadruple in Q_1.
6. The connections in graph $\mathcal{G}$ represent sensors associated to a common image point. Each completely connected sub-graphs, that is, clique, in $\mathcal{G}$ represents the sensors common to the same image point. The number of

maximal cliques provides an estimate of the number of images. Our goal is to identify each maximal clique.

Estimate Number of Images

1. The maximal cliques in $\mathcal{G}$ define the clusters of sensors, and each such clique corresponds to one image point. The number of image points as a result is estimated by the number of maximal cliques.
2. In general, identifying the cliques in any graph is an NP-complete problem (see Bron and Kerbosch [2]). In the applications that we have in mind, such as identifying a room, the total number of microphones used is not that large; therefore, computing the number of maximal cliques for such specific applications may not be too complex.
3. There are many tools available to identify cliques in a graph, and we are not focusing on providing efficient algorithms for the clique problems. In this dissertation, we use the maximalCliques function in MATLAB to find the cliques in $\mathcal{G}$. A MATLAB simulation of finding cliques to estimate image locations is given in Section 5.2.
4. For each maximal clique of sensors, we estimate the image location based on distance estimates as discussed in Section 5.2. Consequently, we can form a set $\mathcal{I}^*$ of estimated image locations, which gives us an estimated room shape.

Image Estimation

1. The set $\mathcal{I}^*$ consists of image estimates based on the quadruples in $\mathcal{Q}_1$. Given that we did not use information from $\mathcal{Q}_2$ and $\mathcal{Q}_3$ to determine $\mathcal{I}^*$,

it is possible that $\mathcal{I}^*$ contains image estimates that are not real images as well as missing images from the original image set $\mathcal{I}$.

2. Further investigation is not provided in this dissertation and could be analyzed for future work.

5.2 MATLAB Simulation

5.2.1 Setup

The simulation setup is as follows:

- The six-walled room is of size $500x300x300$.
- The sound source is at $(0, 0, 0)$ in the center of the room.
- 24 microphones are placed inside the room such that for each image there are four microphones closest to it.
- The distance between the microphone and its closest image is corrupted by $N(0, 1)$ Gaussian noise.

5.2.2 Recovering the Unknown Room

Using $\tau = 2$, we are able to obtain six clusters of sensors by taking five measurements from each sensor, and the resultant image locations are as follows:

$$(-1.1, 299.3, 1.2), (1.1, -301.2, -0.7),$$
$$(498.2, -2.9, -3.2), (-502.6, -3.7, 7),$$
$$(2.8, -5.9, -295.5), (-1.2, -3.8, 302.1).$$

This compares well with the true value of the image locations:

$$(0, 300, 0), (0, -300, 0),$$
$$(500, 0, 0), (-500, 0, 0),$$
$$(0, 0, -300), (0, 0, 300).$$

5.3 Conclusion

In this chapter, we have resolved the problem of identifying an unknown room from noisy wave samples. The physical model for the problem is introduced, and a detailed procedure is proposed. The MATLAB simulation is given at the end of the chapter.

CHAPTER 6

PHYSICAL EXPERIMENT

In this Chapter, we describe experiments with real microphones and speakers to recover the shape of an office room.

6.1 Equipment

The equipment used consisted of microphone array, approximately omnidirectional speaker unit, and a central computing server.

6.1.1 Microphone Array

The microphones that we use in this experiment are identified as Microphone Array Mark III, designed and made by the Information Access Division of National Institute of Standards and Technology. The array consists of two parts: the microphone array board and mother board. The microphone array board samples sound waves, and the mother board provides an interface to computers to control the microphone operations and access the recorded data. Detailed documentation of the microphone array can be found online at `http://www.nist.gov/smartspace/downloads/Microphone_Array_Mark_III_version_2.pdf`. Fig. 6.1 and Fig. 6.2 show the actual microphone arrays that we use.

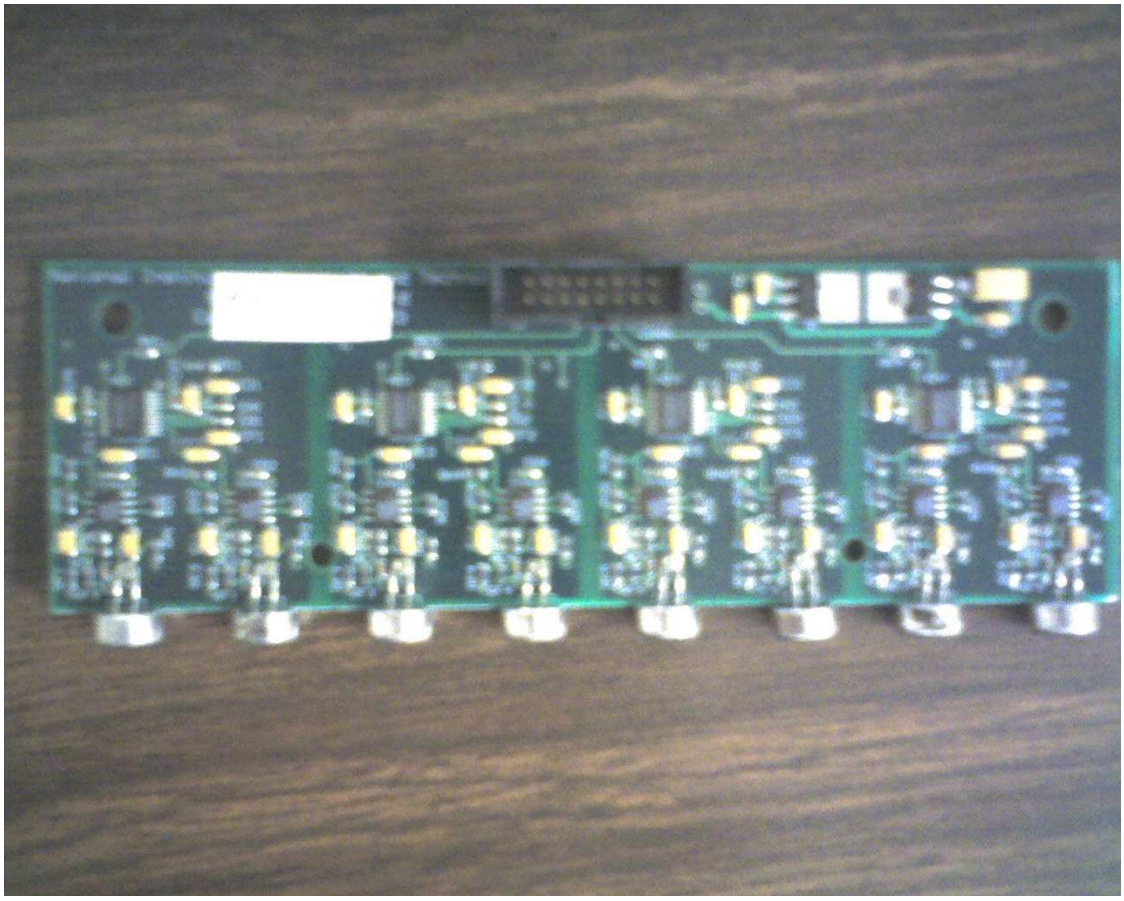


Figure 6.1: Microphone array: Each microphone can be detached from the board by attaching a cable between the microphone and board

6.1.2 Speaker Unit

The speaker unit is constructed by placing four regular computer speakers together, each of which faces one direction (see Figure 6.3). This speaker unit is not omnidirectional in practice, as it only covers the horizontal plane. One of the practical reasons is that connecting together six speakers onto one sound output port in the computer does not provide sufficient power to generate a large-enough sound pulse. In this case, we make a four-speaker unit to approximate the omnidirectional speaker unit.

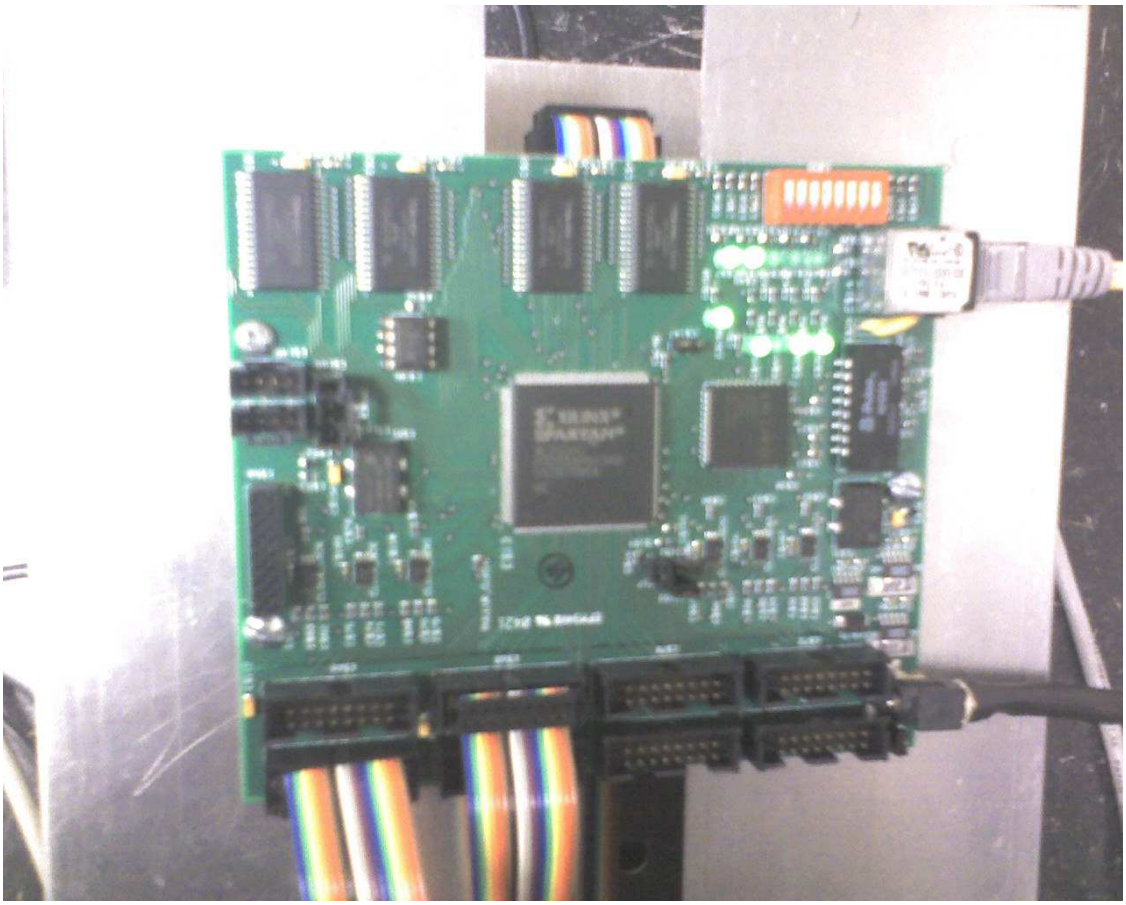


Figure 6.2: Motherboard

6.1.3 Central Computing Unit

A Unix server is used to control the sound from the speaker unit and the recording of the wave by the microphone array (see Figure 6.4).

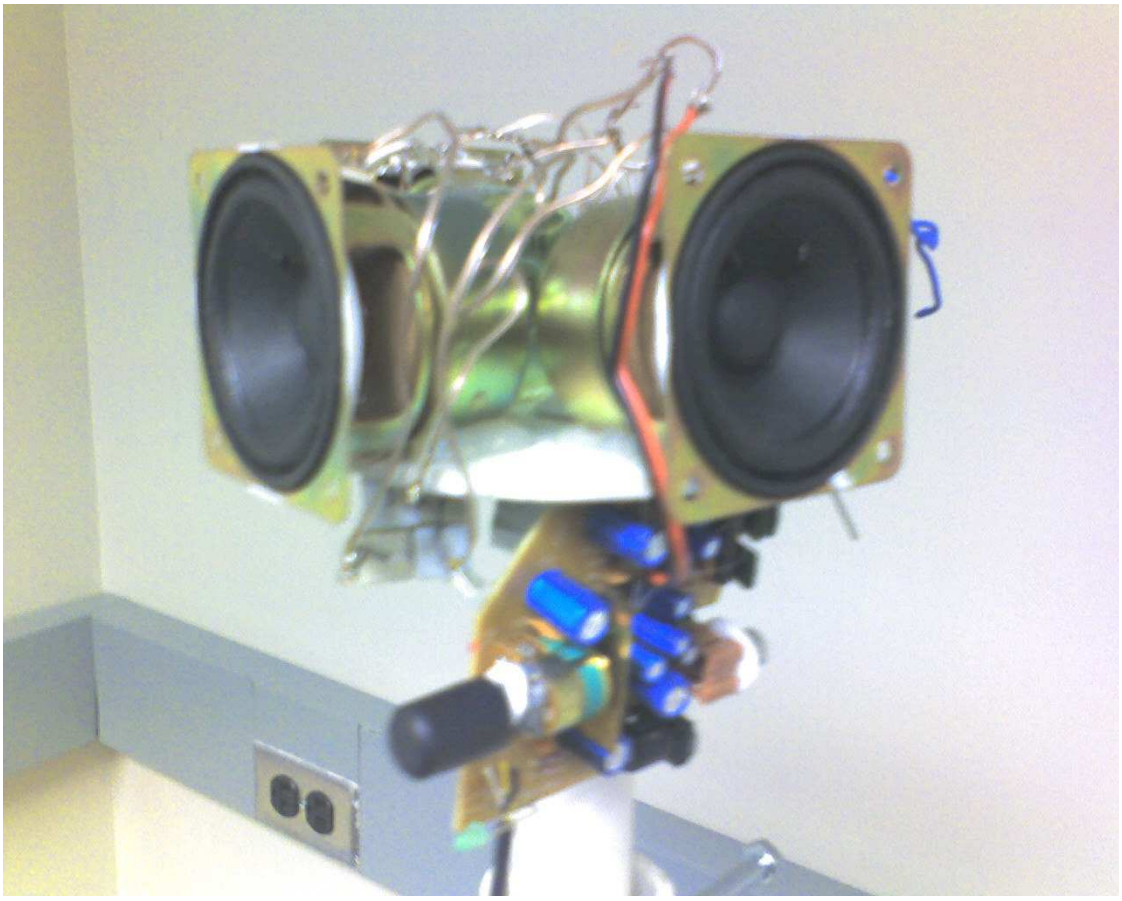


Figure 6.3: Speaker Unit

6.2 Distance Measurement

6.2.1 Hardware Setup

Extracting distances from one microphone and one pulse setup is done as follows:

- The microphone has a sampling rate of 22060 samples/s.
- The speaker is programmed to play a one sample short pulse.
- The microphone is placed 0.8 meters away from a speaker and 0.2 meters

Figure 6.4: The Central Computing Unit

away from the wall.

- The speaker and microphone are on a line that is perpendicular to the wall. The image created by this wall is 1 meter away from the microphone.
- The wave is measured for 0.0073 seconds in time, that is, 160 samples.
- One major noise source in the lab room is from computer fans.

Figure 6.5 for hardware setup illustration:

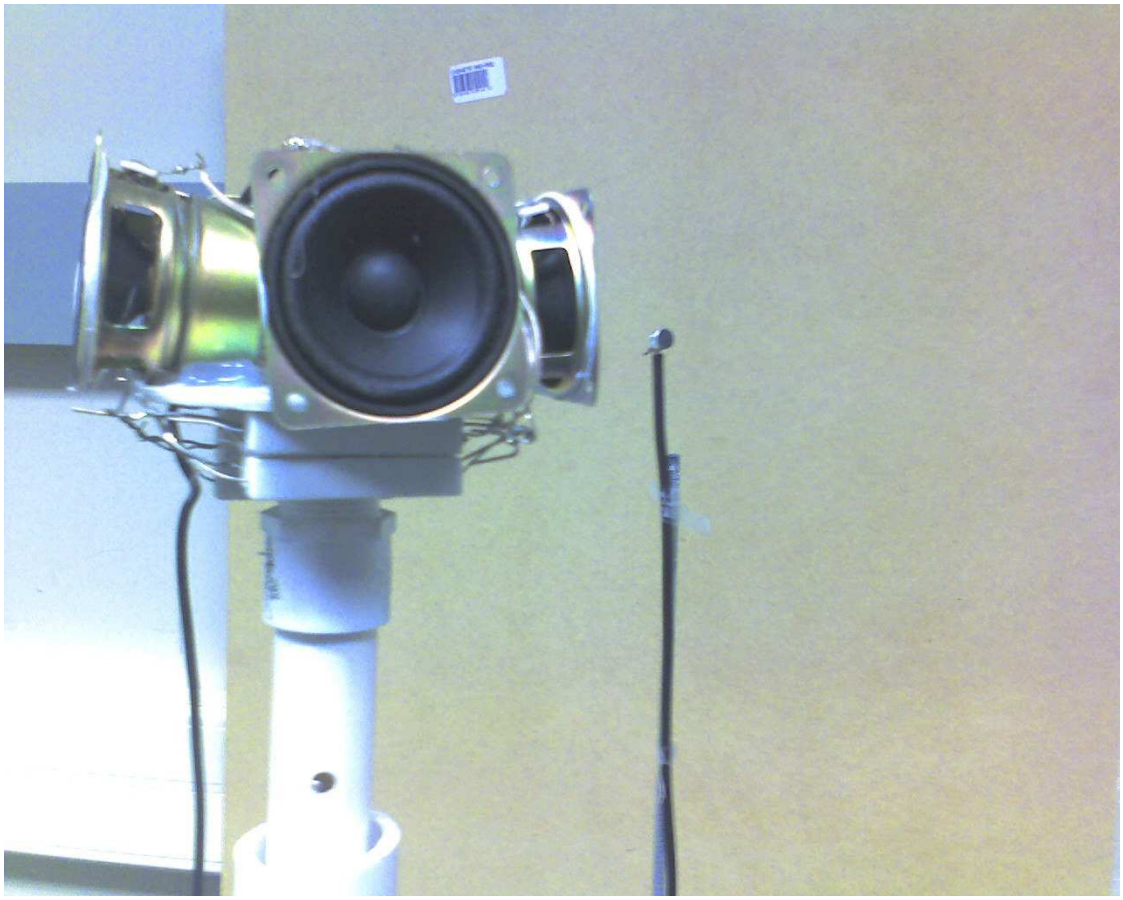


Figure 6.5: Setup for distance measurement experiment

6.2.2 Result

The wave measurement during the 0.0073 seconds is shown in Figure 6.6. Then applying the algorithm in Chapter 2 to extract a distance measurement, we can get the graph of our objective function for which we want to find the minimum. The graph of the objective function $L(\delta)$ is in Figure 6.7. The minimum is at 53 samples, thus $\hat{\delta} = 1.0169$ meters. This results in an error of approximately 0.0169 meters.

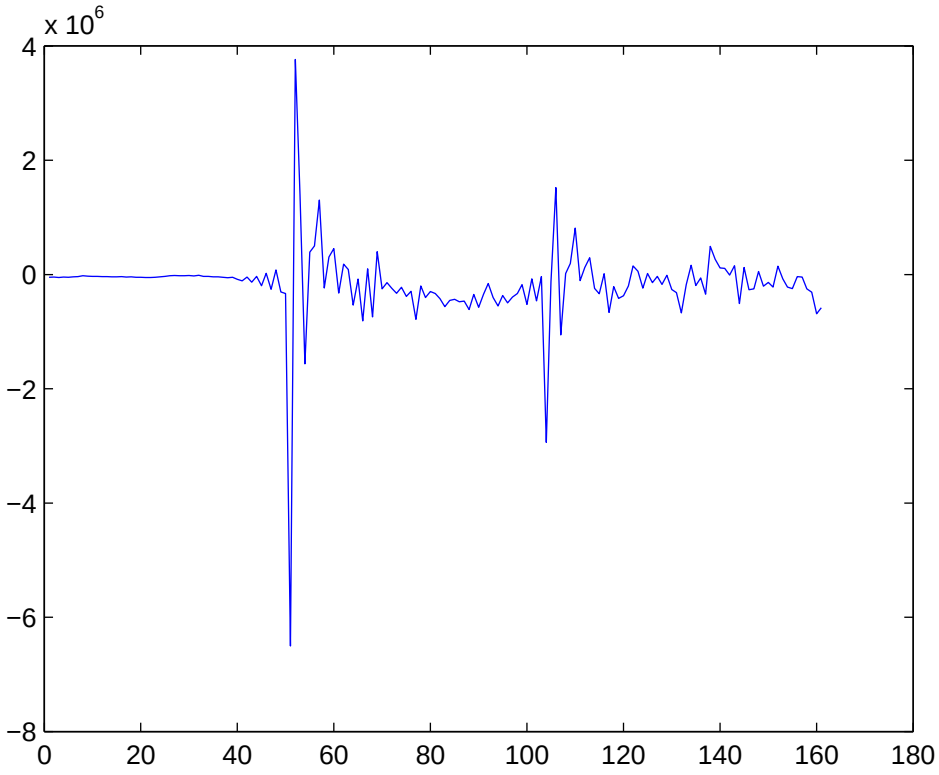


Figure 6.6: The wave measurement in experiment

6.2.3 Analysis

In this real lab experiment, there are several key points to be addressed:

- The error of the estimate in the above experiment cannot be considered very reliable: We physically tape-measured the ”true” distance to determine the error. Tape measurement can be off by 0.01 meters.
- The accuracy of the estimate also depends on the sampling rate: When there is no noise at all, the estimation error may not be zero due to sampling, but the error will be bounded by one sample. This is because if there is no noise, the samples of the first arrival signal and echo signal may dif-

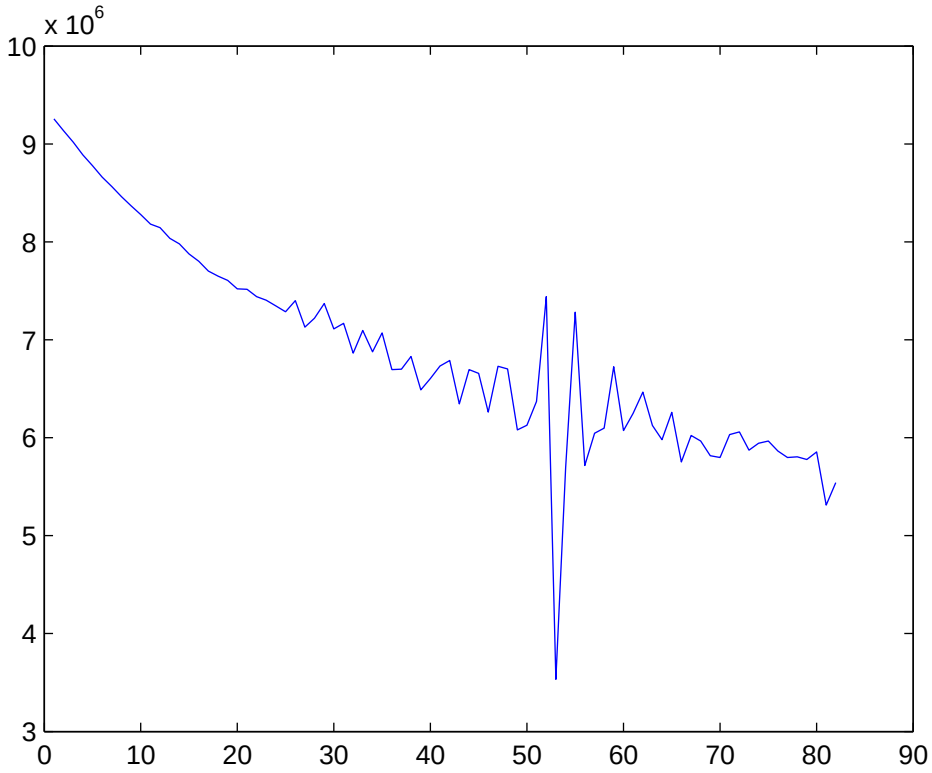


Figure 6.7: The objective function $L(\delta)$

fer only by up to one sample in time. Thus, the estimate will have an error of no more than one sample.

6.3 Recovering a Room

6.3.1 Hardware Setup

Recovering an unknown room from sound wave is tested as follows:

- The room is $500 \times 300 \times 300$cm in size.

- The microphones have a sampling rate of 22060 samples/s.
- The speaker is placed at (0,0,0) and programmed to play a one sample short pulse that is repeated for eight times. The microphones record the sound wave as shown in Figure 6.8.
- The image points are unknown and located at $(500, 0, 0)$, $(-500, 0, 0)$, $(0, 300, 0)$, $(0, -300, 0)$, $(0, 0, 300)$, and $(0, 0, -300)$.
- The microphones are placed in the room so that each image has at least four microphones closest to it.
- One major noise in the lab room is from computer fans.
- As the sound source is not omnidirectional, changing orientation of the speaker set is required during sound sampling.

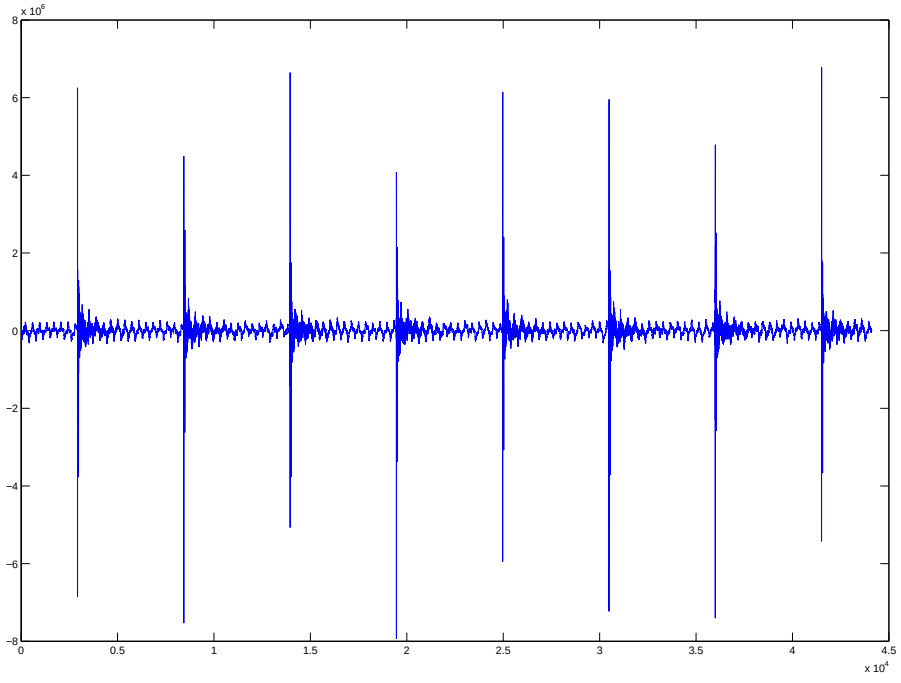


Figure 6.8: Wave measurement at one microphone

6.3.2 Experimental Procedure

1. Place the speaker in the middle of the room, and set the coordinate to (0,0,0).
2. Choose sufficient number of points from the area that speaker unit covers to take wave samples.
3. Change the orientation of the speaker unit, and implement step 2 again.
4. Record the location of the points chosen in steps 2 and 3.

6.3.3 Computational Procedure

1. For each wave sample, we apply the algorithm in Section 3.2.2 to extract distance measurement from each pulse it captured and generate an average distance measurement.
2. Given the locations of each sample point (or microphone), we apply the algorithm from Section 6.1.4 to estimate the number of walls and obtain the estimate of the image locations.

6.3.4 Result

There are a total of 60 measurements taken in the experimental room. And three sets of measurements were chosen to show the results:

1. 30 points of measurements are chosen such that at least four points are closest to each image: The estimated number of image points is exactly six, and the estimate of the image loca-

tions are: (503.26, 8.24, 6.82), (−502.76, 6.34, −4.85), (2.65, 297.52, −3.59), (−6.76, −296.57, −5.76), (3.32, 6.94, 303.84), and (4.35, −8.46, −296.33).

2. 20 points of measurements are chosen randomly out of the 60 points. The estimate of the image points is three. The estimates for the images are (−4.93, 2.75, 308.65), (493.86, 4.21, −0.97), and (7.68, 3.48, −299.45).

6.4 Conclusion

In conclusion, there are two experiments described in this chapter: distance estimation and recovery of an unknown room. Both experiments are conducted under noisy conditions. The distance estimation is very much consistent with what we expected. The recovery of an unknown room has interesting results:

1. Result 1 shows an ideal case, and we are able to recover all the images.
2. Result 2 shows a case in which we fail to provide sufficient number of microphones; therefore, some microphones do not contribute in the estimation, and only partial images are recovered.

CHAPTER 7

CONTRIBUTIONS AND DIRECTIONS FOR FUTURE RESEARCH

7.1 Contributions

There were four key results presented in this dissertation that contribute the novelty. First, we proposed an algorithm to estimate the distance between a microphone and its closest image based on the time of arrival of the acoustic wave. Second, we detailed the identification of an unknown room based on noise-free acoustic measurements. This consisted of problem setup, room identification procedure, and, most importantly, the probabilistic analysis of recovering room shape. Third, we proposed a solution for room identification with noisy measurements. Finally, we performed the room identification in a real lab setup. The following sub-sections provide a summary of the four key results in the dissertation. Please refer to the previous chapters for more details.

7.1.1 Distance Estimation

To estimate the shape of a convex room, the first and foremost step was to get the information of image points via acoustic waves. Let $s(t)$, compactly supported on $[0, T^*]$, be the source signal generated from a speaker inside a room, and this signal is recorded by a single microphone d_1 distance away. The recorded signal picked up by a microphone is:

$$r(t) = \frac{A}{d_1} s(t - \frac{d_1}{c}) + \frac{A}{d_2} s(t - \frac{d_2}{c}) + n(t),$$

where $n(t)$ is the noise. Assume that $r(t)$ has the support $[0, T]$ and it is large enough to contain the echo signal. Let $r_2(t)$ be:

$$r_2(t) = r(t)1_{[\frac{d_1}{c}+T^*,T]}(t).$$

Then the estimated distance between this microphone and its closest image is:

$$\hat{\delta} = \arg\min_{\delta\in\Delta} \|r_2(t) - \frac{A}{\delta + d_1} s(t - \frac{d_1 + \delta}{c})\|,$$

where $\|\cdot\|$ is L^2-norm over $t \in \mathcal{D}$ as in Section 2.2.

$\hat{\delta}$ contributed the information of image locations by which we could further localize the image and determine the shape of a room. The novelty of $\hat{\delta}$ is that it is easy to compute, it is equal to the true distance when there is no noise, and it converges to the true distance when the probe signal is increased in amplitude.

7.1.2 Noise-Free Room Identification

To crystallize the noise-free room identification problem, we proposed a number of assumptions or models in Chapter 1 and Section 3.1; on top of which, we also introduced the notion of generic arrangement of microphones in Section 3.2.5. The generic arrangement of microphones can be summarized as that the image location can be uniquely determined, in the generic case, by four or more non-coplanar microphones; and other cases are not generic. This allowed us to provide a necessary and sufficient condition, that is, having four or more non-coplanar microphones closest to each image source, to recover the room shape. The procedure to identify room shape with no noise was provided in Section 3.3.

The probabilistic analysis of the noise-free room identification was introduced in Section 3.3 by providing distribution models for microphone placement, number of walls, and Voronoi region sets. These allowed us to obtain a closed form result for the probability of recovering the room shape (Section 3.4).

The result of the noise-free room identification contributes a two-fold impact: First, the modeling of the problem provided the platform for us to solve the problem without trivializing it; second, based on our model, we were able to provide a measurable solution to the noise-free room identification problem, that is, a procedure to recover the room shape and a closed-form probabilistic analysis.

7.1.3 Room Identification with Noise

The first step taken to deal with room identification with noise was to estimate the image location based on noisy distance measurements. The solution to this estimation problem was provided in Chapter 4. With this estimate, there was still one major challenge left, that is, to identify which microphones were closest to which image. Unlike in the noise-free case, a method based on Wald's sequential analysis was provided in Section 5.1.2 to cluster microphones to the closest image by which an ad-hoc approach to inferring image points in noisy case was shown in Section 5.1.3. The novelty of the method in Section 5.1.2 was that it did not only provide a way to cluster microphones but also contributed the confidence interval for the clustering. This provided us some measures on the performance of the approach in Section 5.1.3 for room identification with noise.

7.1.4 Physical Experiment

The results of two confirming lab experiments for the room identification were reported in Chapter 6: 1) distance estimation based on noisy measurements, and 2) room shape recovery in noise. Those experiments verified the performance of the algorithms for detecting room shape. They were repeatable, and conducted with off-the-shelf equipment that met the budget and lab space constraints. There was also potential for improvement on the experiments, such as using more powerful speakers to better the experiment results.

7.2 Directions for Future Research

This dissertation assumed that our only prior information about the unknown room was that it was a convex polyhedron, of unknown spatial extent and of unknown number of faces/walls, and that its interior surfaces specularly reflected incident sound waves. Furthermore, our only source of additional information came from a network of spatially distributed microphones driven by a single sound source that emitted a known signal s of short duration T^*. Possible directions for extending our analysis and the problem setup, without radically departing from the research direction taken in this dissertation, include the following:

1. additional prior information that further constrains the room structure, such as that it is also a parallelepiped that may or may not be rectangular or has a known number of faces/walls;
2. weakening the assumption of its being a convex polyhedron but keep-

ing the assumption that it is a polyhedron with specularly reflecting surfaces/walls;

3. further analysis of the use of those microphones that either were silent because they received no echoes within time T and or whose responses to the sound source were ignored because they did not lead to the inclusion of the microphone in a quadruple of microphones associated with the same image point;

4. possible usefulness of an array of sound sources, each emitting a distinct signal.

7.2.1 Rectangular Parallelepiped Hypothesis

We illustrate the above-mentioned Point 1 by a rectangular parallelepiped room shown in Fig 7.1 in which a sound source S is placed inside the room and generates six image points $I_1, I_2, I_3, I_4, I_5, I_6$. Given this prior knowledge, it is not necessary to have four microphones closest to an image point in order to recover its location. Given an image point I_1 recovered through four microphones placed closest to its associated face/wall, we immediately know that the image point I_2 determined by the opposite wall lies on the line passing through S, I_1, as shown in Fig 7.1. Therefore we can generically use two microphones that are closest to I_2 to recover its location: The first identifies two possible points, the intersection of a sphere with the line, and the second, in the generic case, ensures that we choose the correct point for I_2. Furthermore, having identified I_1 and I_2, and knowing that the room is a *rectangular* parallelepiped, the rest of the problem becomes two dimensional: The remaining image points must lie in the plane orthogonal to the line $S\,I_1$ through S. So three microphones are enough for

identifying I_3, and two for I_4 with knowledge of I_3. Then the problem becomes one dimensional, so two microphones are enough for I_5 and two for I_6. Hence, with this additional assumption as to the shape of the unknown room, we can reduce the number of microphones required to recover the room shape from 24 to 15. We leave to future research the exploration of the impact of supplemental assumptions that restrict the possible shapes of the convex polyhedral room.

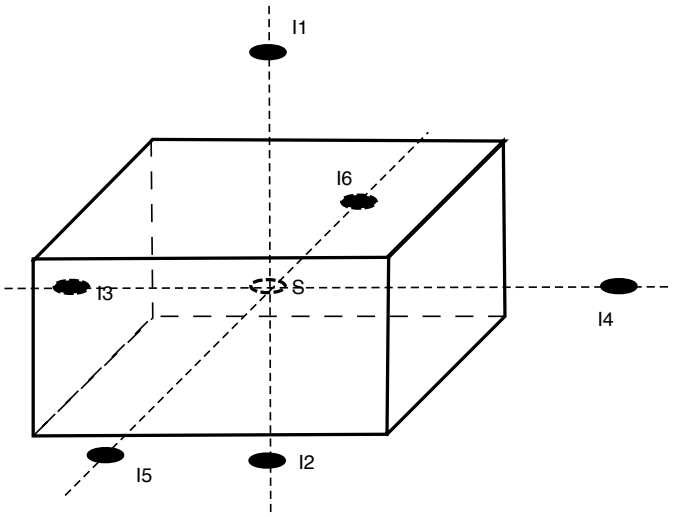


Figure 7.1: Example of using a rectangular cuboid room

7.2.2 Non-Convex Rooms

Pursuant to our second point, if the original room is not convex, then our approach of intersecting half-spaces generated by the image points corresponding to the walls of the non-convex room will not recover the true room—the intersection of half-spaces will always yield a convex room.

We illustrate this by an example in 2D, the lower dimension chosen for the sake of visual clarity. We start with the non-convex room, shown in Fig 7.2,

that is determined by vertices V_1, V_2, V_3, V_4 and walls W_1, W_2, W_3, W_4. Placing a source signal, S, as shown in this room, generates four image points, I_1, I_2, I_3, I_4 corresponding to the four walls. (A different placement of S can lead to the identification of only three image points and only three walls.) Applying our approach of intersecting the half-spaces determined by the image points, we obtain the convex room shown in Fig 7.3 and having vertices at V_1', V_2', V_3', V_4', and walls W_1', W_2', W_3', W_4'. In this example, when we intersect half-spaces we come up with a convex polyhedral room that is a subset of the true non-convex polyhedral room.

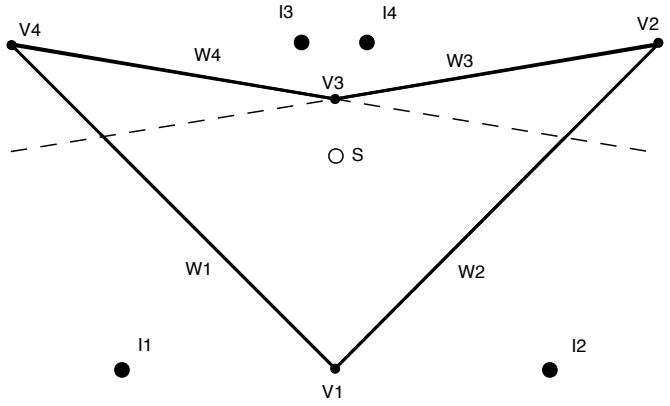


Figure 7.2: Example of a non-convex room

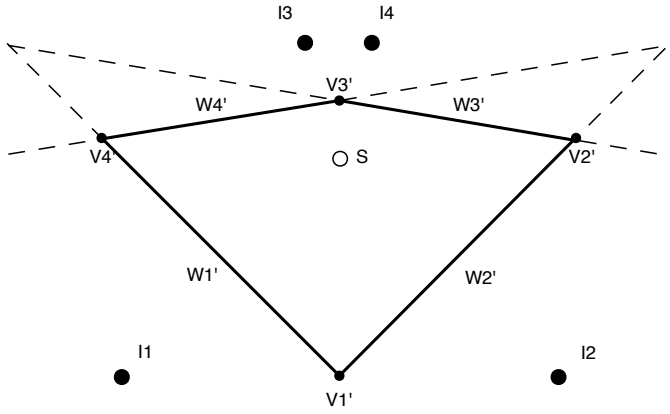


Figure 7.3: Example of recovering a convex room

Can we tell that the room is not convex from the system of measurements we have studied in this dissertation? We notice from Figure 7.4 that if the room is not convex, then we do not always have a direct (line-of-sight) path between the sound source and a microphone. Such an event was not considered in this dissertation. Our arrangement of the microphone array presumed line-of-sight connections between the sound source and each microphone.

Nevertheless, proceeding as illustrated in Fig 7.4, the direct path between sound source S and microphone M is blocked by the corner at vertex V_3 formed by wall W_3 and W_4. As a result, the time of arrival of the first wave at microphone M will not correspond to the time that the wave would have traveled on the direct path. In fact, it will be delayed, as the wave incident on M has to undergo reflection from W2 along the path (S-N-M shown in Fig. 7.4). Consequently, if we observe that the first wave arrival at a microphone is delayed beyond the expected time of arrival of a direct wave, which is known to us from our placement of source and microphones, we can conclude that there is a non-convexity obstructing the line-of-sight path between this microphone and the sound source.

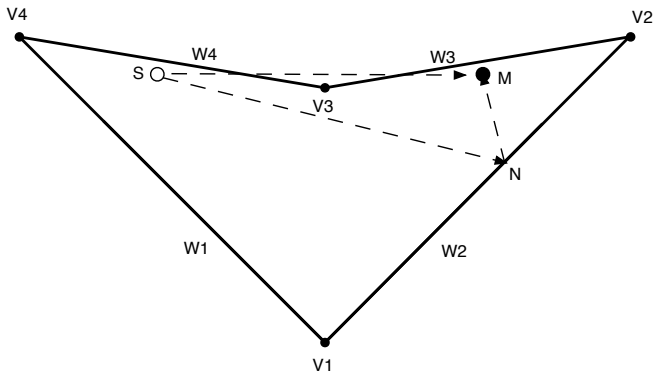


Figure 7.4: Example of non-convexity identification

Perhaps if we suspect the room to be non-convex then, based on the above

examples, we might adopt a policy of exploring the room through moving the sound source to different locations (see 7.2.4). All of these speculations take us well beyond the problem analyzed in this dissertation, and their exploration is left to future research.

7.2.3 Silent Microphones and Ignored Microphones

Regarding Point 3, in this dissertation there were two types of microphones that are not used to determine the shape of the room: *silent* microphones, the ones that did not detect any echo, and *ignored* microphones, the ones that picked up echoes but did not contribute to the determination of any image point.

The existence of silent microphones is due to hardware limitations and our prior ignorance of the room size. If our allowed recording time is T, the greatest distance we can survey is cT where c is the speed of sound. We chose to discard silent microphones if observation to time T did not detect an echo. However, these silent microphones do provide some negative information about the image points. Let $\mathcal{S}$ denote the set of silent microphones. We know that at least for noiseless case:

1. If microphone $m_k \in \mathcal{S}$ then there is no image point within or at distance cT from location $\mathbf{d}_k$, where cT and $\mathbf{d}_k$ are known to us.
2. For any microphone $m_k \in \mathcal{S}$, if we define a set S_k of points no more than cT distance away from $\mathbf{d}_k$, then set $S = \cup_k S_k$ defines a region that contains no image point.

With the above information, we may be able to eliminate image point candi-

dates if they are in region S. In Fig 7.5 we provide a two-dimensional example. There are three microphones at $\mathbf{d}_1, \mathbf{d}_2$ and $\mathbf{d}_3$ being used to identify one image point. Microphones at $\mathbf{d}_1$ and $\mathbf{d}_2$ are active and generate two circles which provide two intersection points located at $\mathbf{I}_1$ and $\mathbf{I}_2$. The third microphone at $\mathbf{d}_3$ is a silent microphone that determines a dotted-lined circle in which no image point is allowed. In our example, intersections at $\mathbf{I}_1$ and $\mathbf{I}_2$ contribute two candidate image points, and point at $\mathbf{I}_2$ is inside the dotted-lined circle, therefore we can conclude that $\mathbf{I}_2$ is not an image, and $\mathbf{I}_1$ is the image point. Using silent microphones might assist in the identification of image points that are not otherwise identified by at least four (in the 3D case) active microphones. However, in the case of real interest, where there is received noise, the use of silent microphones is less direct and their exploitation remains for future research.

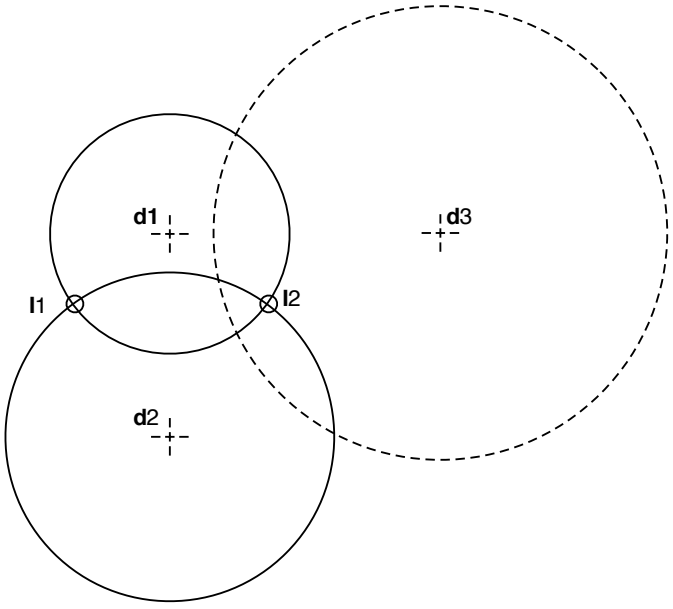


Figure 7.5: Example of using a silent microphone

In addition to silent microphones, this dissertation also ignored certain microphones. The recovery of room shape was dependent upon identifying four

or more microphones that are closest to the same image. As a byproduct of this process, we could have some microphones that did not share a common image point with any other microphones and were therefore ignored. However, at least in the noiseless case, the existence of an ignored microphone provides the information that its closest image lies on a known sphere. Let us form a set $\mathcal{U}$ of ignored microphones whose associated spherical surfaces do not determine any of the recovered image points. In the noiseless case, there are a number of relevant attributes that pertain to the set $\mathcal{U}$:

1. All of the microphones in $\mathcal{U}$ have responded to an echo from an unidentified image point, one that is not recovered.
2. If microphone $m_k \in \mathcal{U}$ then there is an unidentified image point that is precisely at a distance ρ_k from location $\mathbf{d}_k$, where ρ_k and $\mathbf{d}_k$ are known to us.
3. $\mathcal{U}$ provides some information about the number of undiscovered image points. For instance, if there are K microphones in $\mathcal{U}$ whose spherical surfaces are mutually disjoint, then there are at least K undiscovered image points. This gives a lower bound of the undiscovered image points.

At least in the noiseless case, the ignored microphones provide some information that is relevant to the determination of additional image points and, hence, additional walls of the room. It remains for future research to explore how to effectively use the noisy information from ignored microphones to assist in room shape recovery.

7.2.4 Multiple Sound Sources

In this dissertation we focused on imaging the room through an array of microphones responding to a single source. Particularly in the case of non-convex rooms, it can be useful to also deploy an array of sound sources, thereby improving our chances of being able to detect and infer non-convex rooms. It would not be difficult to distinguish these sources not only by their locations but also by their emitted signals, with s_k being the signal from the k-th source. By recording signals received at a microphone so as to include not only a first echo but also a second echo, we would be able to determine which sources produced the two echo signals. This line of inquiry is a promising direction for future research.

BIBLIOGRAPHY

[1] J. B. Allen and D. A. Berkley. Image Method for Efficiently Simulating Small-Room Acoustics. *J. Acoust. Soc. Am.*, 65(4):943–950, 1979.

[2] C. Bron and J. Kerbosch. Algorithm 457: Finding All Cliques of an Undirected Graph. *Communications of the ACM Vol. 16, Issue 9. ACM Press: New York, USA*, 1973.

[3] P. R. Cromwell. Polyhedra. *Cambridge University Press, Cambridge, United Kingdom*, 1997.

[4] J. Elson, L. Girod, and D. Estrin. Fine-Grained Network Time Synchronization using Reference Broadcasts. *In Proc. 5th Symp. Op. Syst. Design Implementation (OSDI)*, Boston, MA, 2002.

[5] T. L. Fine. Probability and Probabilistic Reasoning for Electrical Engineering. *Prentice-Hall/Pearson, Upper Saddle River NJ*, 2006.

[6] W. H. Foy. Position-location solutions by Taylor-series estimation. *IEEE Transactions on Aerospace and Electronic Systems AES-12, 2, p187-194*, March 1976.

[7] C. Gordon, D. L. Webb, and S. Wolpert. One Cannot Hear the Shape of a Drum. *Bulletin of the American Mathematical Society*, 27(1):134–138, 1992.

[8] C.W. Helstrom. Elements of Signal Detection and Estimation. *PTR Prentice Hall*, 1995.

[9] A. Hu and S. D. Servetto. Algorithmic Aspects of the Time Synchronization Problem in Large-Scale Sensor Networks. *ACM/Kluwer Mobile Networks and Applications*, 10:491–503, 2005.

[10] A. Hu and S. D. Servetto. On the Scalability of Cooperative Time Synchronization in Pulse-Connected Networks. *IEEE Trans. Inform. Theory*, 52(6):2725–2748, 2006.

[11] M. Kac. Can One Hear the Shape of a Drum? *Amer. Math. Monthly*, 73(4):1–23, 1966.

[12] D. E. Manolaski. Efficient Solution and Performance Analysis of 3-D Position Estimation by Trilateration. *IEEE Transactions on Aerospace and Electronic Systems vol32, p1239-48*, October 1996.

[13] P. K. Sen. Theory and Applications of Sequential Nonparametrics. *Society for Industrial Mathematics*, 1987.

[14] D. Siegmund. Sequential Analysis: Tests and Confidence Intervals. *Springer*, 1985.

[15] M. I. Skolnik. Introduction to Radar Systems Third Edition. *Tata McGraw-Hill Publishing Company Ltd.*, 2001.

[16] F. Zhao and L. Guibas. Wireless Sensor Networks: An Information Processing Approach. *Morgan Kaufmann*, 2004.

[17] M. Zhao and S. D. Servetto. An Analysis of the Maximum-Likelihood Estimator for Localization Problems. *In the Proceedings of the 2nd IEEE/CreateNet International Workshop on Broadband Advanced Sensor Networks, Boston, MA*, October 2005.